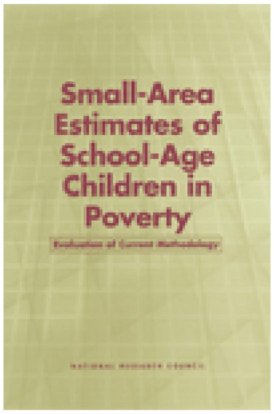


Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology

Panel on Estimates of Poverty for Small Geographic Areas, Constance F. Citro and Graham Kalton, Editors, Committee on National Statistics, National Research Council

ISBN: 0-309-50157-1, 270 pages, 6 x 9, (2000)

Small-Area Estimates of School-Age Children in Poverty

Evaluation of Current Methodology

Panel on Estimates of Poverty for Small Geographic Areas

Constance F. Citro and Graham Kalton, *Editors*

Committee on National Statistics

Commission on Behavioral and Social Sciences and Education

National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C.

NATIONAL ACADEMY PRESS 2101 Constitution Avenue, N.W. Washington, D.C. 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract No. RN96131001 between the National Academy of Sciences and the U. S. Department of Education. Support of the work of the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (Number SBR-9709489). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number 0-309-07301-4

Additional copies of this report are available from National Academy Press, 2101 Constitution Avenue, N.W., Lockbox 285, Washington, D.C. 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu

Suggested citation: National Research Council (2000). *Small-Area Estimates of School-Age Children in Poverty: Evaluation of Current Methodology.* Panel on Estimates of Poverty for Small Geographic Areas, Constance F. Citro and Graham Kalton, editors. Committee on National Statistics. Washington, D.C.: National Academy Press.

Printed in the United States of America

THE NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

PANEL ON ESTIMATES OF POVERTY FOR SMALL GEOGRAPHIC AREAS

GRAHAM KALTON (*Chair*), Westat, Rockville, Maryland
DAVID M. BETSON, Department of Economics, University of Notre Dame
NANCY E. DUNTON, Midwest Research Institute, Kansas City, Missouri
WAYNE A. FULLER, Department of Statistics, Iowa State University
THOMAS B. JABINE, Consultant, Washington, D.C.
SYLVIA T. JOHNSON, School of Education, Howard University
THOMAS A. LOUIS, School of Public Health, University of Minnesota
SALLY C. MORTON, RAND, Santa Monica, California
JEFFREY S. PASSEL, Urban Institute, Washington, D.C.
J.N.K. RAO, Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario
ALLEN L. SCHIRM, Mathematica Policy Research, Inc., Washington, D.C.
PAUL R. VOSS, Department of Rural Sociology, University of Wisconsin
JAMES H. WYCKOFF, Graduate School of Public Affairs, State University of New York, Albany
ALAN M. ZASLAVSKY, Department of Health Care Policy, Harvard Medical School

CONSTANCE F. CITRO, *Study Director*
MICHAEL L. COHEN, *Senior Staff Officer*
MICHELE VER PLOEG, *Research Associate*
MEYER ZITTER, *Consultant*
TELISSIA M. THOMPSON, *Senior Project Assistant*
JAMIE CASEY, *Project Assistant*

COMMITTEE ON NATIONAL STATISTICS
1999-2000

Contents

Preface

The Panel on Estimates of Poverty for Small Geographic Areas was established by the Committee on National Statistics at the National Research Council in response to the Improving America's Schools Act of 1994. That act charged the U.S. Census Bureau to produce updated estimates of poor school-age children every two years for the nation's more than 3,000 counties and 14,000 school districts. The act also charged the panel with determining the appropriateness and reliability of the Bureau's estimates for use in the allocation of more than $7 billion of Title I funds each year for educationally disadvantaged children.

Our charge was both a major one and one with immovable deadlines. The panel had to evaluate the Census Bureau's work on a very tight schedule in order to meet legal requirements for allocation of Title I funds. As it turned out, we produced three interim reports: the first one evaluated county-level estimates of poor school-age children in 1993, the second one assessed a revised set of 1993 county estimates; and the third one covered both county- and school district-level estimates of poor school-age children in 1995. This volume combines and updates these three reports into a single reference volume.

The reference volume is intended to serve two purposes. First, it provides specific documentation of the Census Bureau's current methods for producing small-area estimates of poor school-age children, the evaluations that have been conducted of them to date, and their advantages and limitations for Title I fund allocations. Second, it offers a case study of the development, evaluation, and application of model-dependent small-area estimates that may be helpful for future work in small-area estimation and the use of small-area estimates for such important public policy purposes as fund allocations. As a case study, it makes clear the complexity of the estimation task and the necessity of comprehensive evaluations of the quality of the estimates.

This reference volume is a companion to the panel's final report (National Research Council, 2000). That report outlines an agenda for research and development of the Census Bureau's income and poverty estimates for small areas, including further research and development for the Bureau's current models and the possible uses of new survey and administrative records data sources for improving those models. It also discusses issues about the use of such estimates for public programs.

We could not have carried out our work over the past several years without the cooperation and help of many people. The panel notes, first, the many people in the U.S. Departments of Education and Commerce who contributed to the panel's work of reviewing the Census Bureau's small-area estimates of poor school-age children that are described in this volume. We thank the current and former staff of the Census Bureau who prepared the estimates, many of whom also worked on evaluations of them: David Aultman, William Bell, Patrick Cardiff, John Coder, Robert Fay, Robin Fisher, Matthew Kramer, Esther Miller, Mark Otto, Ronald Prevost, Douglas Sater, Paul Siegel, Cotty Armstrong Smith, Alexander Strand, Jess Thompson, George Train, David Waddington, and Signe Wetrogan. We also thank the Census Bureau staff who facilitated the arrangements for the work: Cynthia Clark, Nancy Gordon, Charles Nelson, and Daniel Weinberg.

Daniel Kasprzyk of the National Center for Education Statistics, who served as project officer for the study for the U.S. Department of Education, was most helpful in facilitating the panel's work throughout the project. The panel also appreciates the help of other Department of Education staff–in particular, Sandy Brown, Thomas Corwin, Lonna Jones, Kay Rigling, William Sonnenberg, and Stephanie Stullich–in providing information and educating us about the allocation process for the Title I program.

The panel also thanks Rona Briere, freelance editor, and Eugenia Grohman, associate director for reports of the Commission on Behavioral and Social Sciences and Education, for helping to combine the text of the panel's three interim reports into a single, seamless volume.

The three interim reports that are combined in this volume were reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the Report Review Committee of the National Research Council. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published volume as sound as possible and to ensure that the volume meets institutional standards for objectivity, evidence, and responsiveness to the

study charge. The review comments and draft manuscripts remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their participation in the review of one or more of the reports that make up this volume: Johnny Blair, Survey Research Center, University of Maryland; James R. Chromy, Statistics Research Division, Research Triangle Institute, Research Triangle Park, NC; Emerson Elliott, National Council for the Accreditation of Teacher Education, Alexandria, VA; Eric Hanushek, Hoover Institution, Stanford University; Robert Hauser, Center for Demography, University of Wisconsin; Lyle V. Jones, L.L. Thurstone Psychometric Laboratory, University of North Carolina; Roderick J.A. Little, Department of Biostatistics, University of Michigan; Lincoln Moses, Department of Biostatistics, Medical Center, Stanford University; William O'Hare, Annie E. Casey Foundation; John Pratt, Graduate School of Business, Harvard University; Nathaniel Schenker, National Center for Health Statistics, U.S. Department of Health and Human Services; Stanley Smith, Bureau of Economics and Business Research, College of Business, University of Florida; Franklin Wilson, Department of Sociology, University of Wisconsin; and Kirk Wolter, National Opinion Research Center, Chicago, IL.

Although the individuals listed above provided constructive comments and suggestions, it must be emphasized that responsibility for the final content of this volume rests entirely with the authoring panel and the institution.

Finally, I thank my fellow panel members and the project staff for all their efforts. The panel members willingly gave their time, commitment, hard work, and good cheer to our endeavor to provide valuable and timely information for allocating funds as fairly as possible for poor school-age children. Constance Citro has performed in a truly outstanding manner as the project's study director, and she has been very ably assisted by Michael Cohen, Michele Ver Ploeg, Meyer Zitter, Telissa Thompson, and Jamie Casey. It has been a pleasure to work with the panel and the project staff over the past four years to carry out our very challenging charge.

Graham Kalton, Chair
Panel on Estimates of Poverty for
Small Geographic Areas

Small-Area Estimates of School-Age Children in Poverty

1

Introduction and Overview

In the early 1990s the U.S. Census Bureau began work to produce regularly updated estimates of key income and poverty measures for subnational areas in a program called SAIPE–Small Area Income and Poverty Estimates. The estimates are produced by using sophisticated statistical modeling techniques with data from multiple sources, including the March Current Population Survey, the 1990 decennial census, and administrative records.

Legislation passed in 1994 called for the use of updated Census Bureau estimates of poor school-age children for counties and school districts to allocate more than $7 billion of funds each year under Title I of the Elementary and Secondary Education Act. The same legislation also authorized the U.S. Department of Education to commission a review of the Census Bureau's estimates by a panel of the National Research Council's Committee on National Statistics. The statute required that the department use the updated Census Bureau estimates unless the Secretaries of Commerce and Education determined that some or all of the estimates are "inappropriate or unreliable" on the basis of the panel's study (Improving America's Schools Act of 1994 [P.L. 103-382] and 1996 continuing resolution).

The Panel on Estimates of Poverty for Small Geographic Areas was set up to carry out the authorized study. The panel was charged with a broad review of the Census Bureau's SAIPE model-based estimates for small geographic areas and their utility for fund allocations and other purposes.

The panel began its work in June 1996 and produced three interim reports. Each report evaluated a specific set of estimates of poor school-age children from the Census Bureau and made recommendations about their use for Title I allocations: 1993 county estimates (National Research Council, 1997); revised 1993 county estimates (National Research Council, 1998); and 1995 county and school district estimates (National Research Council, 1999). This report combines in a single reference document the information in the panel's three interim reports about these estimates: what they are, how they were produced, and their quality. The panel's final report (National Research Council, 2000) provides an agenda for research and development for the Census Bureau's SAIPE Program in the next decade. It covers modifications to the Bureau's current models; possible uses of new sources of data from surveys and administrative records for improving the models, and issues in using the model-based estimates for such program purposes as fund allocation.

The panel hopes that this reference report will be useful for people who require small-area estimates of poor school-age children and for people with an interest in the methods and applications of small-area poverty estimates more broadly. Since the development of model-based or model-dependent estimates that combine data from multiple sources is a complex task,[1] such estimates should always be accompanied by complete documentation of how they were developed and a full evaluation of their quality. The Census Bureau has carried out evaluations and prepared documentation, and this report also contributes to that goal.

TITLE I

Title I supports compensatory education programs to meet the needs of educationally disadvantaged children (see Moskowitz et al., 1993). From the enactment of the program in 1965 through the 1998-1999 school year, the role of the Department of Education has been to allocate funds to the nation's more than 3,000 counties (including Puerto Rico as a county equivalent), and the states have then distributed the county allocations to school districts. For the 1999-2000 school year, in response to the Improving America's Schools Act of 1994, the department for the first time made allocations directly to almost 15,000 school districts (formally known as local educational agencies, LEAs).[2]

The Title I allocations use estimates of formula-eligible children: predominantly poor school-age children, who are defined by the Census Bureau to be

[1]By "model-dependent" we mean that the accuracy of the estimates depends on the validity of the assumptions of the estimation model.

[2]The intent of the Title I legislation, when it was originally enacted in 1965, was that the Department of Education would allocate funds directly to school districts; however, lack of data with which to develop school district estimates led to the two-stage allocation system that was used through the 1998-1999 school year.

related children aged 5-17 in families with incomes below the poverty level.[3] (Related children include family members under age 18 in a household, except married sons, daughters, or spouse of the householder and foster children.) Historically, the allocations made by the Department of Education to counties used the estimates of poor school-age children from the most recent decennial census for which data were available. The estimates from one census were used for a decade or more until estimates from the next census became available. Since the proportions and numbers of children in poverty can change significantly over time, the 1994 legislation called for the use of updated estimates of poor school-age children for Title I allocations. The Census Bureau was to provide updated estimates for counties in 1996, for use in the Title I allocations to counties in the 1997-1998 and 1998-1999 school years, and then to provide estimates for school districts in 1998 and every 2 years thereafter, for use in direct Title I allocations to school districts in the 1999-2000 and later school years. Having the most up-to-date estimates possible is important so that resources can be directed towards areas that are most in need.

At present, Title I funds are provided for two different types of allocations–basic grants and concentration grants. Under the two-stage allocation process used through the 1998-1999 school year, basic grants were provided to all counties and suballocated to school districts that had at least 10 formula-eligible children and whose percentage of formula-eligible children exceeded 2 percent of the district's total school-age children. Concentration grants were provided to counties that had high numbers (more than 6,500) or high proportions (more than 15%) of formula-eligible children and suballocated to eligible school districts in those counties. However, with the direct allocation process first used for the 1999-2000 school year, the provisions for county eligibility and grant amounts no longer apply; concentration grants are now provided directly on the basis of school district eligibility.

Under the direct allocation process, the department determines the initial allocation amounts for all school districts. However, a provision in the 1994 legislation permits states to reallocate these amounts for school districts with less than 20,000 population by using another data source that the department approves. This provision was included because of concerns about the likely quality

[3]The poverty status of individuals is determined by comparing the before-tax money income of their families to the appropriate poverty threshold. The poverty thresholds vary by family size and are updated by the change in the Consumer Price Index each year. See National Research Council (1995a) for an evaluation of the current official poverty measure and a proposed alternative measure; the issue of how poverty should be defined is not considered in this volume. The Title I allocations also take account of the average per-pupil expenditures in each state, as well as the allocations made in the previous year (through a "hold-harmless" provision).

of estimates of poor school-age children for small school districts. About four-fifths of school districts contain fewer than 20,000 people, although these districts contain only about 27 percent of all school-age children in the United States. For the 1999-2000 and 2000-2001 school year allocations, nine states used this option.

UPDATED ESTIMATES

County Estimates

The Census Bureau was initially charged to produce updated estimates of poor school-age children at the county level for use in Title I allocations for the 1997-1998 school year. For this purpose, the Census Bureau provided county estimates of the numbers of school-age children in 1994 in families with incomes below the poverty level in 1993. The estimates were developed from a statistical regression model that used administrative data from Internal Revenue Service and Food Stamp Program records for 1993, estimates of poor school-age children in 1989 from the 1990 census, and 1994 population estimates to predict county numbers of poor school-age children in 1993 as measured in the March Income Supplement to the Current Population Survey (CPS).

The model was estimated for counties with one or more households with poor school-age children in the CPS sample–about one-third of total counties. To increase the reliability of the predictions, the model used weighted averages of 3 years of data from the March 1993, 1994, and 1995 CPS, covering income in 1992, 1993, and 1994. For counties in the CPS sample, the model predictions were combined with the direct CPS 3-year average estimates for those counties, in a procedure that weighted the two estimates according to their relative precision. For the remaining counties (two-thirds of the total), the model prediction for a county was the estimate for that county. As a last step, the estimates from the county model were calibrated to estimates from a similar statistical model for states.

The data used in the county model are obtained from several sources, and most data are not available until 2 years after the period to which they refer. When the developmental work began in 1994, the Census Bureau decided that it would not be able to produce estimates in time for the 1997-1998 allocations for a later year than 1993, given the time required for acquiring, processing, and applying the data for a new statistical model.

In its first interim report (National Research Council, 1997), the panel reviewed the Census Bureau's modeling approach favorably but concluded that there had not been sufficient time to thoroughly evaluate the updated estimates produced by the specific model that the Bureau developed. As an interim solution for Title I allocations for the 1997-1998 school year, the panel recommended that the 1993 county estimates be averaged with 1990 census estimates. This recommendation was adopted. Subsequently, the Census Bureau completed an extensive evaluation of the county model, modified it in several respects, and produced a revised set of 1993 county estimates of poor school-age children. In its second interim report (National Research Council, 1998), the panel recommended that the revised 1993 county estimates be used for Title I allocations for the 1998-1999 school year, which was done.

For both the 1997-1998 and 1998-1999 school years, the Department of Education used the Census Bureau's poverty estimates to make allocations to counties. As in the past, the states then allocated the county amounts to school districts. The states used a variety of data sources for these allocations: many states used 1990 census data wholly or in part; some states used such data sources as numbers of children approved to receive free or reduced-price lunches under the National School Lunch Program or children in families receiving Aid to Families with Dependent Children (or its successor program, Temporary Assistance to Needy Families) in each district. For basic (but not concentration) grants in some states in which the boundaries of school districts bore little relationship to county boundaries, the department permitted the state to ignore the county allocations in dividing up the total allocation amount for the state among school districts. The Department of Education must approve a state's allocation plan but is not required to approve the specific estimates used by a state or the allocation amounts.

School District Estimates

For Title I allocations for the 1999-2000 school year, the Census Bureau was charged to provide the Department of Education with updated estimates of poor school-age children for school districts. The 1994 legislation required the department, in turn, to make direct allocations to school districts rather than to counties unless the Secretaries of Education and Commerce determined that the school district estimates were inappropriate or unreliable for this purpose, taking into account the panel's recommendations. Under this procedure, the states would not be involved, unless they elected to exercise the provision in the 1994 legislation that permits a state to reallocate the Department of Education's allocation amounts for all school districts in the state that have an estimated 20,000 or fewer people.

There appear to be several reasons that Congress in the 1994 legislation deemed it desirable for the Department of Education to make direct allocations to school districts. First, direct allocations by the department impose a measure of consistency on the allocation process. Second, direct allocations to school districts solve a problem with the concentration grant formula in which a county may not be eligible for a concentration grant, but one or more of the school districts in the county may meet the eligibility criteria. (This can happen when a poor school district is located within a county that, on average, is not poor enough to qualify.) Under a two-stage allocation process, poor school districts in counties that do not qualify for a concentration grant would receive less funds than they would receive with direct allocations.[4] Finally, if adequate data were available for estimation, the use of updated school district-level estimates in the allocations would take account of changes that have occurred since the previous census in poverty among school districts within counties.

In early 1999 the Census Bureau provided estimates for school districts of the numbers of school-age children in 1996 who were living in families with incomes below the poverty level in 1995. Developing reliable updated estimates for counties is not easy, and the task is much more difficult for school districts. Some school districts are the same as counties. However, most school districts are smaller than counties, many of their boundaries cross county lines, and the boundaries can and often do change over time. Also, some school districts provide education for specific grade levels, such as K-8 or 9-12. Largely because of these complicating factors, there is a paucity of data for developing updated poverty estimates at the school-district level: there are currently no school district equivalents of the Internal Revenue Service or Food Stamp Program data that are used in the Census Bureau's state and county estimation models.

Because of the lack of data at the school district level, the Census Bureau's procedure for developing 1995 school district poverty estimates used a simple model that assumes that the proportions or shares of poor school-age children in school districts within each county in 1995 were the same as they were in 1989 (as measured by the 1990 census). The estimation procedure involved the following steps: 1990 census data were retabulated to match 1995-1996 school district boundaries (determined from a special survey); the proportion of the county total of poor school-age children in the 1990 census was determined for each school district (or part of a school district) in the county; and the 1990-based proportions were then applied to updated 1995 county estimates from the Census Bureau's county model to produce 1995 school district estimates.

Because of the time required to complete the survey of 1995-1996 school district boundaries and the time lags in the availability of data for the county model, the Census Bureau was not able to produce school district estimates for later than 1995 to be used in allocations for the 1999-2000 school year. Moreover, the Census Bureau's shares-based estimation procedure did not capture intracounty variation in the extent to which school-age poverty increased or decreased among school districts between 1989 and 1995. In addition, the estimates of school district shares of poor school-age children within counties based on 1990 census long-form data were subject to high levels of variability due to sampling error for many small districts. However, the estimation procedure produced estimates more recent than the census, it was consistent across the

[4] States could reserve up to 2 percent of their concentration grant funds to allocate to such districts.

nation, and it responded to the concern that concentration grants be directed to all eligible school districts, including those in counties that were not eligible.

In its third interim report (National Research Council, 1999), the panel concluded that although the Census Bureau's 1995 estimates of poor school-age children had potentially large errors for many school districts, the estimates were nonetheless not inappropriate or unreliable to use for direct Title I allocations to districts as intended by the 1994 legislation. In reaching this conclusion, the panel interpreted "inappropriate and unreliable" in a relative sense. Some set of estimates must be used to distribute Title I funds to school districts. The panel concluded that the Census Bureau's estimates were generally as good as–and, in some instances, better than–estimates that were previously used. On the basis of the panel's study, the Department of Education made direct allocations to school districts for the 1999-2000 and 2000-2001 school years by using the Census Bureau's 1995 school district estimates and other elements of the allocation formula. The department also notified the states of a recommendation by the panel that states electing to reallocate amounts for school districts with fewer than 20,000 people on the basis of some other data source (e.g., school lunch data) should do so on a county-by-county basis so as to reflect (approximately) the Census Bureau's updated estimates of poor school-age children from the county model.

PLAN OF THE REPORT

This reference volume describes and evaluates the Census Bureau's methodology for producing estimates of poor school-age children for counties and states for 1993 and 1995 and for school districts for 1995.[5] The report brings together material in the panel's three interim reports to provide a comprehensive description of the current estimation methodology and evaluation results. Similar methods, with likely small modifications, will be used by the Census Bureau to produce state, county, and school district estimates of poor school-age children for the immediate future and to produce other SAIPE poverty estimates, which include total numbers of poor people and poor people under age 18 for states and counties and, for states only, numbers of poor children under age 5. (The SAIPE Program also produces median household income estimates for states and counties.) In the longer run, research and development of data sources and estimation methods will likely lead to changes in the methodology for improved estimates (see National Research Council, 2000).

This reference report contains nine chapters and five appendices. Chapter 2

[5]These estimates are available on the Census Bureau's web site: http://www.census.gov/hhes/www/saipe.html.

describes key features of the Title I allocation formula, such as hold-harmless provisions and thresholds for eligibility, that affect the kinds of estimates that are required and their application and evaluation. The chapter also describes the two-stage allocation process that was used for the 1998-1999 school year and the direct allocation process that was used for the 1999-2000 school year.

Chapter 3 describes and compares the input data sources that the Census Bureau uses to develop state and county estimates of poor school-age children. These sources include the decennial census, the March Current Population Survey, tax return data, and Food Stamp Program data. (Another source is population estimates from the Census Bureau's postcensal population estimates program, which are described in Chapter 8.) The chapter also reviews trends in poverty over time.

Chapters 4-6 describe and evaluate the Census Bureau's procedure for obtaining updated county estimates of the numbers and proportions of poor school-age children in 1993 and 1995. Chapter 4 describes the 1995 county and state models and differences from the 1993 models; Chapter 5 describes alternative 1993 county models that were evaluated; and Chapter 6 provides evaluation results for the 1995 and 1993 models. Although the Department of Education does not use county estimates in Title I allocations when the allocations are made directly to school districts, the county estimates are central to the method used by the Census Bureau to derive updated school district estimates and, therefore, to an evaluation of those estimates. The state estimates of poor school-age children are described and evaluated as well because they are used in deriving the county estimates.

Chapter 7 describes and evaluates, as best as can be done, the data and procedures the Census Bureau used to develop 1995 school district estimates of poor school-age children. Given the scarcity of data with which to implement alternative estimation procedures for school districts, the opportunities for evaluation are very limited.

Chapter 8 describes and evaluates the Census Bureau's procedure for obtaining, from its population estimates program, state and county estimates of the total number of school-age children for 1994 and 1996 and school district estimates of the total number of school-age children and the total population in each district for 1996.

Chapter 9 outlines research and development activities for further work on developing updated county and school district estimates of poor school-age children in the near term (see also National Research Council, 2000:Ch.3).

The appendices cover the following topics: models for county and state poverty estimates (A); regression diagnostics on alternative county regression models (B); county model comparisons with 1990 census estimates (C); use of National School Lunch Program data in New York State to estimate school-age children in poverty for school districts (D); and the estimation procedure for Puerto Rico, which is treated as a county and school district equivalent in the Title I allocation process (E).

2

Title I Allocation Procedures

This chapter summarizes the procedures used to allocate Title I funds, describing features that need to be considered when evaluating the reliability and appropriateness of the Census Bureau's SAIPE estimates of poor school-age children for use in the allocations. Following a summary description of the Title I formulas, the chapter describes the two-stage procedure that was used for county and school district allocations from the inception of the program in 1965 through the 1998-1999 school year and the direct allocation procedure that was first used for school district allocations for the 1999-2000 school year.[1]

TITLE I FORMULAS

Title I allocations are based on estimates of formula-eligible children, which comprise four groups: related poor school-age children, as estimated by the Census Bureau; children in foster homes; children in families above the poverty level that receive Temporary Assistance to Needy Families (TANF);[2] and children in local institutions for neglected and delinquent children. The Census Bureau's estimates of poor school-age children account for about 95 percent of total formula-eligible children.

[1] School districts are also known as local educational agencies (LEAs).
[2] The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 abolished Aid to Families with Dependent Children (AFDC) and replaced it with TANF.

The statute contains four formulas for allocating Title I funds–basic grants, concentration grants, targeted grants, and the Education Finance Incentive Program–but Congress has to date appropriated funds only for the basic and concentration formulas. Basic grants have existed since the program began in 1965; concentration grants were added in 1978 to provide additional funds to school districts with high concentrations of school-age children in poverty (Moskowitz et al., 1993). The total amount of Title I funds allocated for the 1998-1999 school year was \$7.3 billion–\$6.2 billion for basic grants (85% of the total) and \$1.1 billion for concentration grants; the total amount allocated for the 2000-2001 school year was \$7.7 billion (U.S. Department of Education, 2000).

The basic grant formula has a low threshold for eligibility for school districts to receive funds: eligible districts must have at least 10 formula-eligible children and the number of eligible children must exceed 2 percent of the district's population aged 5-17. There were no minimum eligibility criteria for a county to receive a basic grant under the two-stage procedure.

In contrast, the concentration grant formula has a high eligibility threshold, allocating funds only to jurisdictions–counties and school districts for the two-stage procedure, school districts for the direct procedure–with high numbers or high percentages of poor school-age children. To be eligible to receive a concentration grant, a jurisdiction must have more than 6,500 formula-eligible children or more than 15 percent of the children in the jurisdiction must be formula eligible. Under the two-stage procedure, a school district could meet the eligibility threshold for a concentration grant but not receive funds because the county in which it is located was not eligible. States could reserve up to 2 percent of their concentration grants funds to allocate to such districts.

The two formulas take account not only of numbers of formula-eligible children in each jurisdiction, but also of each state's average per-pupil expenditure, a factor intended to compensate for state differences in the cost of education. A state minimum grant provision applies to each of the two formulas as well. For basic grants, the state minimum grant is equal to the lesser of (1) 0.25 percent of total funds available for Title I basic grants and (2) the average of 0.25 percent of total funds and 150 percent of the national average grant payment per formula-eligible child multiplied by the number of formula-eligible children in the state.[3] There is some added complexity for the state minimum for concentration grants (Moskowitz et al., 1993).

Finally, the Title I formulas include hold-harmless provisions to cushion the impact of decreases in allocations. Historically, there was no hold-harmless provision for concentration grants, but for school year 1996-1997, the Title I

[3]The national average grant payment for basic grants is the total amount of basic grant funds divided by the number of formula-eligible children in school districts that are eligible for basic grants.

legislation specified a hold-harmless provision for both formulas at a rate of 100 percent: that is, a jurisdiction that met the eligibility threshold could not receive fewer funds than it had received in the previous school year. For school year 1997-1998 and beyond, the legislation specified for basic grants that the hold-harmless provision was to be applied at variable rates, with a higher rate for higher poverty jurisdictions: those with 30 percent or more poor school-age children were to be guaranteed at least 95 percent of the prior year's basic grant; the guarantee was to be 90 percent for those with 15-30 percent poor school-age children and 85 percent for those with fewer than 15 percent poor school-age children. For school year 1997-1998 and beyond, the legislation did not include a hold-harmless provision for concentration grants.

However, for allocations for school years 1998-1999, 1999-2000, and 2000-2001, Congress enacted a 100 percent hold-harmless provision for both basic and concentration grants. In addition, for school years 1999-2000 and 2000-2001, Congress stipulated that school districts that were no longer eligible for a concentration grant under the new direct allocation procedure using the Census Bureau's 1995 estimates of poor school-age children would nonetheless receive 100 percent of the previous year's concentration grant. Previously, a district had to meet the eligibility threshold (more than 6,500 or more than 15% formula-eligible children) to receive concentration grant funds. Districts continued to have to meet the much lower eligibility threshold to receive a basic grant.

TWO-STAGE ALLOCATIONS

Under the two-stage procedure used for Title I allocations through the 1998-1999 school year, the U.S. Department of Education determined the allocation amounts for each county, and the states then suballocated these amounts to the school districts in their state. The Department of Education calculated basic grant allocations in an iterative process. The number of formula-eligible children in each county was multiplied by 40 percent of the state's per-pupil expenditure;[4] the resulting allocations were then proportionally reduced so that the total matched the total appropriation for basic grants. Allocations were then adjusted to meet hold-harmless and state minimum grant provisions (see above). The department calculated concentration grant allocations on the basis of numbers and proportions of formula-eligible children, the state's per-pupil expenditure, a state minimum provision, and a hold-harmless provision.

States used a variety of data sources for determining suballocations of Title I funds to school districts, as listed below for school year 1997-1998 from a chart provided by the department. The Department of Education approved each state's

[4]For this calculation the state per-pupil expenditure was set to 80 percent or 120 percent of the national average if it fell below or above these limits.

allocation plan but not the specific estimates used by a state or the allocation amounts.

- Seven states and the District of Columbia made no suballocations to districts because their school districts are coterminous with counties (three of these states made suballocations to a few districts in their states that are not coterminous with counties, such as a city that is a separate district from the remainder of the county).
- Eight states used 1990 census data alone.
- Ten states used 1990 census data and estimates of the other categories of formula-eligible children, such as foster children.
- Nine states used a combination of 1990 census data together with counts of children approved to receive free meals or free or reduced-price meals under the National School Lunch Program, or counts of children in families receiving AFDC, or a composite of AFDC, food stamps, and Medicaid data.
- Eight states used free lunch data only.
- Three states used free and reduced-price lunch data.
- One state used free lunch and state tax information.
- Three states used AFDC data only or in combination with foster child data.
- One state used food stamp data.

Over time, it appears that a growing number of states elected to use free (or free and reduced-price) school lunch data to distribute the county allocations to school districts. Interviews conducted by panel staff in early 1999 determined that this trend was continuing: several states that had used AFDC or food stamp data for suballocations were found to have switched to using school lunch counts, although such data include children with family incomes above the poverty threshold. Children with family incomes up to 130 percent of the poverty threshold may receive a free school lunch; children with family incomes between 130 percent and 185 percent of the poverty threshold may receive a reduced-price lunch.

Under the two-stage procedure, most states were constrained to suballocate county amounts to the school districts (or parts of school districts) within each county. However, the Department of Education permitted nine states to make direct allocations of basic grants–but not concentration grants–to school districts without regard to the county allocation amounts because so many of their school districts crossed county boundaries. Of these nine states, one used 1990 census data to make direct allocations of basic grants; five used 1990 census data and estimates of the other categories of formula-eligible children; one used a combination of 1990 census and free and reduced-price lunch data; one used free lunch data; and one used free and reduced-price lunch data.

DIRECT ALLOCATIONS TO SCHOOL DISTRICTS

The Department of Education made direct allocations of Title I funds to school districts for the first time in spring 1999 for the 1999-2000 school year. As directed by the "Improving America's Schools Act" of 1994, the Department based these allocations on the Census Bureau's updated estimates of the numbers of poor school-age children at the school district level (which the panel determined were not unreliable or inappropriate for this purpose when compared with other estimates–see National Research Council, 1999) together with estimates of the other groups of formula-eligible children for school districts that it obtained from the U.S. Department of Health and Human Services and the states. In some cases, the department had to prorate county totals for these other groups to districts when no district-level information was available.

The Census Bureau provided three sets of school district estimates: (1) estimates of school-age children (aged 5-17) in 1996 who were living in and related to a family in poverty in 1995;[5] (2) estimates of all school-age children; and (3) estimates of the total population of the district. The first two sets of estimates were needed to implement the allocation formulas for both basic and concentration grants. The third set of estimates was needed to identify school districts with fewer than 20,000 people. These school districts had to be identified because the 1994 act provided that states, at their discretion, could aggregate the fund allocations for districts with less than 20,000 population and redistribute the funds by using another method approved by the Department of Education.

Nine states applied and were granted approval to reallocate the department's allocations for school districts with less than 20,000 population. All nine states had previously used data sources other than the decennial census, or in combination with the census, to suballocate the department's county allocations under the old two-stage procedure. The data sources the nine states used for reallocating funds for small school districts included:

- Two states used the 1995 SAIPE estimates (weighted at one-half) and another source (weighted at one-half), either counts of children in families receiving public assistance in the state or free lunch counts;
- One state used the 1995 SAIPE estimates (weighted at three-fourths) and free and reduced-price lunch counts (weighted at one-fourth);
- One state used free lunch counts (weighted at one-half) and poor children estimated from state income tax data (weighted at one-half);
- One state used free lunch counts;
- One state used free lunch and free milk counts from public and nonpublic schools;

[5] See Chapter 1 for the definition of "related children."

- Two states used free and reduced-price lunch counts;
- One state used the 1995 SAIPE estimates (weighted at 0.155), foster children counts (weighted at 0.155), free lunch counts (weighted at 0.46), and reduced-price lunch counts (weighted at 0.23).

All nine states used counts of other categories of formula-eligible children, such as children in locally operated institutions for neglected children, in addition to the sources listed above.

3

Data Sources for County Estimates

This chapter describes the data sources used for the Census Bureau's SAIPE Program model-based estimates of poor school-age children for counties for 1993 and 1995. Data sources used for estimates of poor school-age children for school districts for 1995 are discussed in Chapter 7.

The data sources reviewed below are used not only to produce the initial county estimates from the county regression model, but also to produce the state estimates to which the initial county estimates are controlled (see Chapter 4). These sources include the March Current Population Survey (CPS), which provides the dependent variable in the state and county regression models, and the 1990 decennial census, Food Stamp Program administrative records, and federal income tax return administrative records, which provide predictor variables for the state and county models. The state and county regression models also use population estimates from the Census Bureau's postcensal population estimates program, which are described in Chapter 8.

The CPS income estimates that are used to form the dependent variable in the state regression model pertain to the estimation year–data from the March 1994 CPS for income year 1993 for the 1993 SAIPE state estimates; data from the March 1996 CPS for income year 1995 for the 1995 SAIPE state estimates. The county regression model uses an average of 3 years of CPS data, centered on the estimation year–data from the March 1993, 1994, and 1995 CPS for income years 1992, 1993, and 1994 for the 1993 SAIPE county estimates; data from the March 1995, 1996, and 1997 CPS for income years 1994, 1995, and 1996 for the 1995 SAIPE county estimates. The food stamp and IRS data used in the models pertain approximately to the estimation year; the 1990 census data are for income year 1989.

Prior to the introduction of the SAIPE estimates, the census was the sole basis of poverty estimates for Title I allocations. The SAIPE county and state estimates used in recent allocations derive from CPS-based models that reflect a somewhat different standard of measurement than the census. The discussion therefore reviews differences between decennial census and CPS estimates of poverty.

CENSUS DATA

Traditionally, the decennial census has been the source of estimates of poor school-age children for counties, with each census being used for Title I allocations until data from the next census became available. Many states also used census data for suballocations to school districts (see Chapter 2). The 1980 census data, covering income and poverty for 1979, were used for Title I county allocations for the 1983-1984 through 1993-1994 school years (and, in part, for the 1982-1983 school year). The 1990 census data, covering income and poverty for 1989, were used for county allocations for the 1994-1995 through 1996-1997 school years, and were averaged with the 1993 SAIPE county estimates for allocations for the 1997-1998 school year.

In the 1990 census, income data–the basis for measuring poverty–were collected in the long-form sample survey. The long form includes the small number of items that are asked of every household on the short form and other questions that are unique to the long form. The long-form sample in 1990 was about 15.7 million households, or about 1 in 6 households spread systematically across the country, except that very small counties and places (with estimated 1988 populations under 2,500) were sampled at a 1-in-2 rate, and very populous census tracts (or equivalent areas) were sampled at a 1-in-8 rate.

Data in the census are collected mainly by self-enumeration, in which respondents fill out questionnaires received in the mail. In 1990 approximately 70 percent of households that received the long-form questionnaire returned their questionnaires with some or all of the requested information; return rates were somewhat higher (75%) for households that received the short-form questionnaire (National Research Council, 1995b:189-190). Data from the balance of the population were obtained by enumerators who interviewed a household member or, failing that, a neighbor or landlord. The enumerators were mainly inexperienced temporary workers who were given limited training.

The income data in the 1990 census are based on seven questions on various components of income, such as wages and salaries and Social Security benefits. The long form also included a total income question, which was intended to permit respondents to enter a single amount if they could not provide amounts by source. Nonresponse rates are higher for income than for most other items in the census. When household income information is missing, the Census Bureau uses statistical techniques to impute it on the basis of nearby households with similar characteristics. For the 1990 census, on average, 19 percent of aggregate household income was imputed (National Research Council, 1995b:387).

All censuses are subject to undercount–that is, failure to count everyone. There is also overcount, when persons are counted more than once or when ineligible persons are counted. For 1990, the net undercount (gross overcount minus gross undercount) was estimated at 1.8 percent for the total population, but there were substantial differences among population groups categorized by race, ethnicity, and age. Minorities were more heavily undercounted than others. By age, almost two-thirds of the estimated omitted population consisted of two groups: children under age 10 and men aged 25-39 (Robinson et al., 1993:13). The undercount was higher in large cities than in other areas and was disproportionately concentrated in the inner areas of those cities. These are also the areas where poverty is high. There are no direct estimates of the undercount for poor school-age children. However, it seems likely that the undercount for poor school-age children is larger than the undercount of all school-age children.

Decennial census data on income are estimates, and as such they are subject to sampling error because the data are collected from only a sample of households. Although sampling errors are relatively small for large geographic areas, such as states, the sampling errors for smaller geographic areas can be large relative to the estimate.

Table 3-1 provides information on the amount of error due to sampling variability in the estimated numbers of poor school-age children by county from the 1990 census. For example, for 63 counties, the margin of error due to sampling variability is less than 5 percent of the estimated number of poor school-age children.[1] The estimates for these counties are thus fairly precise. Moreover, these counties, although a small percentage (2%) of all 3,141 counties in 1990, are large ones: they contained 37 percent of the nation's poor school-age children estimated by the 1990 census. However, for 1,405 counties, the margin of error due to sampling variability is 25 percent or more of the estimated number of poor school-age children. Although these counties contained only 6.4 percent of the poor school-age children in the nation estimated by the 1990 census, the imprecision in their estimates is of concern for Title I allocations.

---

[1]The margin of error is expressed in Table 3-1 as the relative width of the 90 percent confidence interval; that is the width of the interval as a percentage of the estimated number. Confidence intervals for a sample estimate are ranges that include the average result of all possible samples with a known probability; they are constructed from the estimate and its standard error (the measure of the magnitude of sampling variability of the estimate). The 90 percent confidence interval for an estimate is from 1.645 standard errors below the estimate to 1.645 standard errors above the estimate: there is a 90 percent chance that the 90 percent confidence interval includes the average estimate from all possible samples.

TABLE 3-1 Distribution of Counties by Relative Widths of the 90 Percent Confidence Interval for the Estimated Number of Poor Related Children Aged 5-17 in 1989: 1990 Census

Relative Width of Confidence Interval[a]	Counties		Poor Children	
	Number	Percent	Number	Percent
All Counties	3,138	100.0	7,544,737	100.0
Less than 5%	63	2.0	2,818,997	37.4
5 to 10%	236	7.5	1,846,546	24.5
10 to 15%	466	14.9	1,258,897	16.7
15 to 20%	538	17.1	761,149	10.1
20 to 25%	430	13.7	372,733	4.9
25 to 50%	1,061	33.8	449,464	6.0
50 to 75%	238	7.6	31,585	0.4
More than 75%	106	3.4	5,366	(Z)[b]

NOTE: Three counties with no poor related children aged 5-17 in the sampled households are excluded from the table.

[a]The relative width of the confidence interval is the percentage that the width of the 90 percent confidence interval represents of the estimated number of poor related children aged 5-17 in a county. The 90 percent confidence interval is 3.29 times the standard error of the estimate.
[b]Less than .05 percent

SOURCE: Data from U.S. Census Bureau.

Because the census is taken only once every 10 years, the data do not reflect current socioeconomic conditions and demographic distributions in the population. Concerns about using outdated decennial census poverty estimates for Title I allocations were reinforced by changes observed between the 1980 and 1990 censuses. Nationally, the number of poor school-age children rose by 5 percent over the 10-year period, from 7.7 million to 8.1 million. At the state level, there was considerable variability: 24 states and the District of Columbia experienced declines in the number of poor school-age children of up to 34 percent; 15 states saw increases of up to 25 percent; 8 states had increases ranging between 25 and 50 percent; and 3 states had increases between 50 and 67 percent (Moskowitz et al., 1993:71).

When considering the use of 1990 census data for allocations to be made later in that decade and into the next decade, there were similar concerns about the use of out-of-date information. Income data collected in the 1990 census are referenced to 1989; they do not reflect either the recession that began in 1990 or the recovery that began in 1991 and, consequently, do not reflect changes in the proportion and geographic distribution of people below the poverty level that resulted from the rise and subsequent decline in unemployment and related economic and demographic changes. The belief that census data would not accurately reflect changes in need over time and across areas was a prime impetus for developing and using SAIPE estimates that reflected more up-to-date information. Although the level of poverty for an area at one point in time may not be a good measure of the area's poverty level at another point in time, there is nonetheless a relationship between the two measures. The county and state regression models take advantage of this relationship by including 1990 census poverty levels as predictor variables for estimating poverty later in the decade.

CPS DATA

The CPS is designed primarily to provide monthly estimates of labor force participation, employment, and unemployment. Every March, the CPS collects additional data on income for the prior calendar year from which poverty rates can be determined. The CPS is therefore a more timely source of data on poverty than the census. Indeed, the annual March Income Supplement to the CPS provides the official national measure of poverty.[2] The March Income Supplement also serves as a basis for some federal fund allocations (U.S. Office of Management and Budget, 1993).

For the period from 1990 to 1994, the CPS sample included about 60,000 households each month that were eligible for interview; starting in 1996, this sample size was reduced to about 50,000 households each month. Of eligible occupied households, about 94-95 percent provide an interview. To obtain more reliable income data for the Hispanic population, all November CPS households with one or more Hispanic persons are reinterviewed in March if they still include a Hispanic person. This procedure adds about 2,500 Hispanic households to the sample in March.[3]

The CPS sample design, which is a multistage probability sample design, is revised about once every 10 years on the basis of the results of the latest census.

---

[2]The Survey of Income and Program Participation (SIPP) is another source of up-to-date income and poverty data. Two Committee on National Statistics panels have recommended that SIPP become the official source of annual national poverty estimates in place of the March CPS (see National Research Council, 1993, 1995a).

[3]Beginning in 2001, the size of the sample that is asked the CPS income supplement questions will almost double compared to the current size of the March CPS. Part of this expansion will occur by increasing the monthly CPS sample size in selected states, and part will occur by interviewing a subset of households in the February and April CPS samples and a subset of households that were formerly in the CPS sample. This initiative is being implemented to respond to a congressional mandate for reliable estimates by state of low-income children who lack health insurance coverage.

From 1986 to 1994, the CPS sample design included 727 sample areas consisting of about 1,300 counties. These areas were chosen on the basis of 1980 census data to represent the noninstitutional population in all 3,141 counties (in 1990) and independent cities in the 50 states and the District of Columbia. A design based on the 1990 census was phased in between April 1994 and July 1995: it included 792 sample areas consisting of about 1,300 counties, chosen to represent all 3,143 counties (in 1994) and independent cities in the 50 states and the District of Columbia. In January 1996, the number of sample areas was reduced from 792 to 754.

In general, larger states have larger CPS sample sizes. The largest states, however, have CPS sample sizes that are smaller than their proportionate share of the U.S. population, and the smallest states have proportionately larger sample sizes. For example, California, with 12.2 percent of the U.S. population, has 9.9 percent of the CPS sample; Wyoming, with 0.18 percent of the U.S. population, has 1.3 percent of the CPS sample. This sample design means that estimates of numbers and proportions poor in large states are generally more precise than those in smaller states. The largest states, however, have larger relative errors due to sampling variability than would be expected if the CPS sample were allocated to the states in proportion to their population; the reverse holds true for smaller states.[4]

The CPS is carried out by permanent, experienced, and well-trained interviewers, who interview each household the first month it is in the sample in person, with subsequent interviews by telephone.[5] For the March Income Supplement, the CPS asks household respondents about their money income received during the previous year, using a detailed set of questions for identifying about 28 different sources. About 20 percent of aggregate household income (about the same percentage as in the census) is imputed–that is, the data are missing and therefore constructed from information from similar households (National Research Council, 1993:Table 3-6).

Like other household surveys, the CPS exhibits population undercoverage at higher rates than the census itself. The coverage ratios for the CPS show the magnitude of the population undercoverage relative to population control totals that update the previous census and are produced by the Census Bureau's population estimates program. Coverage ratios are defined as the estimated survey

[4]To meet national-level reliability criteria for the unemployment rate, the sample size in a few large states (e.g., California, Florida, New York, Texas) is somewhat greater than what would be required by a state-based design. A full description of the CPS design is provided by U.S. Census Bureau and Bureau of Labor Statistics (2000); see also the joint Bureau of Labor Statistics and Census Bureau CPS web site: www.bls.census.gov/cps/mdocmain.html.

[5]Part of the CPS sample is changed each month in a rotation plan: each sampled address is in the survey for 4 months, out of the survey for 8 months, and in the survey for another 4 months. Interviews are conducted for the household found at the sampled address each month.

population before ratio-adjustment to census-based population controls divided by the census-based population controls. (Beginning with the March 1994 CPS, the population controls reflect an adjustment for the undercount in the census.) For March 1994, the ratio of the CPS estimated population to the adjusted population control total (all ages) was 92 percent; for the age group 0-14 years and the age group 15-19 years, the ratios were 94 percent and 88 percent, respectively (U.S. Census Bureau, 1996:Table D-2).

CPS undercoverage is corrected by ratio adjustments to the survey weights that bring the CPS estimates of population in line with updated national population controls by age, race, sex, and Hispanic origin. However, the ratio adjustments do not correct for other characteristics on which the undercovered population might be expected to differ from the covered population. For example, the ratio adjustments reweight equally the sample households within an age-race-sex-Hispanic origin category, when research suggests it is likely that lower income households within a category are more poorly covered than higher income households (see National Research Council, 1985:App.5.1).

The CPS sample size is not large enough to produce detailed information on the changes that occur over time in the geographical distribution of the population in poverty, but the survey can provide some useful indicators. It can illustrate how large changes can occur over short periods of time and how different areas can experience substantially different rates of change. As an example, consider the changes in the distribution in the number of poor people of all ages between 1990 and 1994 (income in 1989 and 1993). The CPS sample is sufficiently large to estimate such changes for 11 states, although the estimates are subject to large sampling errors; see Table 3-2.[6] Overall, the estimated total number of poor people in the country increased by 24.5 percent, but with a wide range across states: 52 percent for Florida and 44 percent for California, but only 7 percent for Illinois and only 4 percent for Texas. Statistical sampling error affects the precision of these estimates, but it is still clear that there were changes over the period and that they differed among states.

The CPS data, when grouped by selected categories of counties and averaged over 3 years to improve precision, show similar changes in the estimated number of poor school-age children, which increased for the nation as a whole by 19.6 percent between 1989 and 1993; see Table 3-3. The increase is evident for counties in all regions of the country, in metropolitan and nonmetropolitan areas, and in all population size categories, but there is substantial variation in the size of the increase. The largest increases are for counties with a population size in the category of 1 million or more (33.1%), other (noncentral) counties in metro-

[6]For these 11 states, the sample was designed to meet reliability requirements for consecutive monthly changes in the unemployment rate.

TABLE 3-2 Change in the Total Estimated Number of Poor Persons between 1989 and 1993 for Selected States: March CPS Data

State	Change in Poverty (in percent)	Standard Error of Estimate (in percentage points)[a]
United States	24.5	1.8
Florida	52.3	12.8
California	44.2	9.4
New Jersey	35.8	15.3
Ohio	27.2	12.3
Pennsylvania	26.8	12.1
New York	26.1	8.7
North Carolina	23.0	11.4
Massachusetts	19.6	13.9
Michigan	19.0	10.6
Illinois	7.3	10.1
Texas	4.1	8.0

[a]3.29 times the standard error gives the 90 percent confidence interval.

SOURCE: Data from U.S. Census Bureau.

politan areas (30.1%), and counties in the West region (30.1%). The smallest increases are for counties with a population size in the category of 10,000-49,999 (3.3%, not statistically significantly different from zero), counties in nonmetropolitan areas (10.8%), and counties in the Midwest region (13.3%).[7]

Although the CPS provides more current data than the decennial census, its much smaller sample size limits its ability to produce estimates for smaller areas. For all but a few very large counties, the CPS sample size is too small to produce reliable estimates. In fact, there is no CPS sample in over one-half of U.S. counties; only about 1,300 counties of 3,143 counties (in 1994) are represented in the sample. And for those counties for which CPS sample data are available, the

[7]The increases in the number of poor school-age children between 1989 and 1993 are the result of increases in the number of school-age children, as well as of increases in the poverty rate for this group. Consequently, for the United States as a whole, the poverty rate for school-age children increased by less than the increase in the number of poor school-age children (11.1% versus 19.6%). The increase in the poverty rate for school-age children, like the increase in their number, varied across regions of the country and types of counties.

TABLE 3-3 Estimated Number of Related Children Aged 5-17 in Poverty by Selected Categories of Counties: 1989 and 1993, March CPS Data

County Category	Children in Poverty, Income Year 1989[a]	Children in Poverty, Income Year 1993[b]	Change in Poverty between 1989 and 1993 (in percent)
U.S. Total	8,036,000	9,613,000	19.6*
Metropolitan			
Central	5,608,000	6,853,000	22.2*
Other	362,000	471,000	30.1*
Nonmetropolitan	2,066,000	2,289,000	10.8*
Region[c]			
Northeast	1,312,000	1,636,000	24.7*
Midwest	1,754,000	1,986,000	13.3*
South	3,296,000	3,813,000	15.7*
West	1,674,000	2,178,000	30.1*
Population Size			
Under 9,999	202,000	243,000	20.3
10,000-49,999	1,489,000	1,538,000	3.3
50,000-99,999	759,000	927,000	22.2*
100,000-499,999	2,143,000	2,448,000	14.2*
500,000-999,999	1,229,000	1,510,000	22.9*
1 million and over	2,214,000	2,947,000	33.1*

*Statistically significant difference from 0 using a 10 percent significance level.

[a]The estimates are 3-year centered averages. For 1989 estimates, averages of March 1989, 1990, and 1991 CPS data were used (reported income in 1988, 1989, and 1990, with population controls derived from the 1980 census).

[b]The estimates are 3-year centered averages. For 1993 estimates, averages of March 1993, 1994, and 1995 CPS data were used (reported income in 1992, 1993, and 1994, with population controls derived from the 1990 census, including an adjustment for the estimated undercount beginning with the March 1994 CPS).

[c]The Census Bureau regions are as follows: Northeast—Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania; Midwest—Ohio, Indiana, Illinois, Michigan, Wisconsin, Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, and Kansas; South—Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, and Texas; West—Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada, Washington, Oregon, California, Alaska, and Hawaii.

SOURCE: Data from U.S. Census Bureau.

estimates of poverty and of the population aged 5-17 are, as a rule, extremely imprecise because of small sample sizes. However, as discussed in Chapter 4, a model-based approach that combines CPS estimates with administrative data in a statistical model can be used to yield estimates for counties that are more up to date than census estimates and have acceptable prediction errors. The Census Bureau's county-level model increases the CPS sample size for counties by combining 3 years of estimates.[8]

DIFFERENCES BETWEEN CENSUS AND CPS DATA

The census and the CPS differ in other ways besides sample size. Even for a census year, the decennial census and the CPS produce different results with regard to children in poverty. Table 3-4 shows the differences between the 1990 census (1989 income) estimates of the number of poor school-age children and the 1989 CPS estimates for the nation as a whole and for various subcategories of counties. Table 3-5 provides a similar comparison of poverty rates. The CPS estimates in the two tables are averages of income data for 1988, 1989, and 1990; averaging is used to improve precision given the small CPS sample size in smaller areas.

Overall, for the U.S. population, the CPS provides an estimate of the number of poor school-age children that is 6.5 percent higher than the decennial census.[9] For most groups of counties, the CPS estimate is also higher than the census estimate, and there is a suggestion of a pattern in which the ratio of the CPS estimate to the census estimate of poor school-age children in 1989 may increase as a function of county size. The panel conducted an analysis to determine whether there were statistically significant differences among the CPS-census

[8]By combining 3 years of data from the March 1993, 1994, and 1995 CPS to produce estimates for 1993, the number of counties represented in the sample increases from about 1,300 to about 1,500. A new 1990 census-based sample design was introduced beginning in the April 1994 CPS; some counties are included in both the new design and the old (1980 census-based) design, but other counties are included in only one design. The average number of sample households for counties represented in one or more of the 3 years is 113; for counties with populations under 10,000, the average number of sample households is 28, and for counties with 500,000 or more people, the average number of sample households is 701. However, several hundred (mostly small) counties with CPS sample households lack any sample households with poor school-age children (see Coder et al., 1996:Tables 1,3).

[9]Some portion of the differences shown for the United States and various kinds of subnational areas may be due to the use of 3-year centered averages for the CPS-based estimates, which included a year (1990 from the March 1991 CPS) in which the poverty rate for school-age children was higher than in either 1989 or 1988. The difference between the 1990 census and the single-year March 1990 CPS in the number of poor school-age children for the United States in 1989 is 4.9 percent, compared with 6.5 percent for the 3-year average figure.

TABLE 3-4 Census and March CPS Estimates of Related Children Aged 5-17 in Poverty in 1989, by Selected Categories of Counties

County Category	Children in Poverty, 1990 Census	Children in Poverty, March CPS[a]	Percentage Difference: CPS – Census as Percent of Census
U.S. Total	7,545,000	8,036,000	6.5*
Metropolitan			
Central	5,021,000	5,608,000	11.7*
Other	347,000	362,000	4.4*
Nonmetropolitan	2,177,000	2,066,000	–5.1*
Region[b]			
Northeast	1,180,000	1,312,000	11.2*
Midwest	1,641,000	1,754,000	6.8*
South	3,174,000	3,296,000	3.9*
West	1,550,000	1,674,000	8.0*
Population Size			
Under 9,999	197,000	202,000	2.5
10,000-49,999	1,489,000	1,489,000	0
50,000-99,999	843,000	759,000	–9.9*
100,000-499,999	1,990,000	2,143,000	7.7*
500,000-999,999	1,124,000	1,229,000	9.3*
1 million and over	1,901,000	2,214,000	16.5*

*Statistically significant difference from 0 using a 10 percent significance level.

[a]The CPS estimates are 3-year centered averages of the March 1989, 1990, and 1991 CPS data (reported income in 1988, 1989, and 1990, with population controls derived from the 1980 census).

[b]See Table 3-3 for the states in each region.

SOURCE: Data from U.S. Census Bureau.

ratios for counties grouped by population size and other characteristics, but did not find such differences. However, this work was very preliminary.[10]

Though not fully researched and understood, differences between census and CPS estimates of poverty may result from the different ways the income data are obtained. The census and CPS use the same official poverty thresholds to deter-

[10]Table 3-4 indicates that the differences between the CPS and census estimates of poor school-age children in 1989 are statistically significant (i.e., significantly different from 0) for all county groups except those with small sample sizes. This finding is not surprising given the large national difference in the two estimates; however, it does not support a conclusion that differences between the ratios of CPS estimates to census estimates are statistically significant across county groups. A different comparison would be needed to establish such differences.

TABLE 3-5 Census and March CPS Estimates of Poverty Rates for Related Children Aged 5-17 in 1989, by Selected Categories of Counties

County Category	Children in Poverty (percent) 1990 Census	March CPS[a]	Difference Between Rates
U.S. Total	17.0	18.0	1.0*
Metropolitan			
Central	16.4	17.9	1.5*
Other	11.4	12.5	1.1
Nonmetropolitan	20.4	19.9	–0.5
Region[b]			
Northeast	14.3	15.5	1.2*
Midwest	14.9	15.8	0.9*
South	20.5	21.3	0.8
West	16.2	17.3	1.1*
Population Size			
Under 2,500	22.9	22.1	–0.8
2,500-4,999	22.2	14.6	–7.6
5,000-9,999	23.1	24.7	1.6
10,000-49,999	20.6	20.9	0.3
50,000-99,999	16.6	15.7	–0.9
100,000-499,999	14.7	15.7	1.0*
500,000-999,999	14.6	15.6	1.0
1,000,000 and over	19.1	21.5	2.4*

*Statistically significant difference from 0 at the 10 percent significance level.

[a]The CPS estimates are 3-year centered averages of data from the 1989, 1990, and 1991 March CPS (reported income in 1988, 1989, and 1990, with population controls derived from the 1980 census).

[b]See Table 3-3 for the states in each region.

SOURCE: Data from U.S. Census Bureau.

mine poverty status,[11] income is counted in both as annual money income received in the previous calendar year, and both are intended to measure the same kinds of income. However, the CPS questionnaire asks respondents to provide income amounts for many more detailed categories than does the census questionnaire. For example, the 1990 census asked respondents to provide a com-

[11]For example, for a family of four the 1999 (weighted average) poverty threshold level was $17,029.

bined income amount for Supplemental Security Income (SSI), Aid to Families with Dependent Children (AFDC), and other public assistance or public welfare payments; the CPS asks separately for SSI, AFDC, and other public assistance or public welfare (including the source). Methodological research suggests that more detailed questions elicit more complete income reports (see National Research Council, 1995a:402-405); however, the extent to which questionnaire differences affect the responses in the CPS and the census is not known.[12]

The CPS and the census also use somewhat different rules for defining the universe to which poverty applies. For example, the CPS includes students living in college dormitories as family members in their parental households; the census considers the dormitory the place of residence and excludes residents of college dormitories from the poverty universe. The result is that somewhat more families with college students may be estimated as living in poverty in the CPS than in the census because a college student in a family increases its size and therefore its poverty threshold but likely does not add appreciably to its income.

The way the data are collected may also result in differences. In the CPS, data are collected through personal contacts (mostly by telephone) made by trained field representatives. In contrast, the census primarily relies on respondents to complete and return a questionnaire by mail. These and other differences imply that CPS-based estimates of poor school-age children represent a somewhat different standard of measurement from decennial census estimates. Consequently, moving from the decennial census to the CPS as the basis for estimates of poor school-age children may have had an effect on the time series of allocations for the Title I program.

ADMINISTRATIVE RECORDS DATA

In addition to predictor variables that are formed from the 1990 census and demographic population estimates, the SAIPE state and county regression models include predictor variables that are formed from administrative records data. Criteria for selecting administrative data sources for this purpose were that the data relate to poverty and that they be available for all states and counties on a consistent basis–that is, obtained using the same definitions and procedures and bearing a similar relationship to poverty across areas. The Census Bureau examined a variety of administrative records and selected two sources as most nearly meeting these criteria: administrative data on recipients of food stamps and federal income tax return reports of child exemptions in families with re-

[12]Another difference is that the 1990 census questionnaire, but not the March CPS questionnaire, included a "total income" question. The intent of this question was to permit respondents to enter a single amount if they could not provide amounts by source.

ported income below the poverty threshold. Neither of these two data sources gives the number of school-age children in poverty as measured by the March CPS, but this is not a problem for model-based estimation: it is necessary only that the variables chosen to be used in the model can provide good predictions of that number.

Food Stamps

The total number of recipients of food stamps is available monthly for states and annually for counties. Eligibility requirements for the program are generally uniform across all states, with some exceptions for Alaska and Hawaii. Two key eligibility requirements are that households must have gross income before deductions that is below 130 percent of the applicable poverty guideline and net countable income that is below 100 percent of the applicable guideline.[13] The gross and net income limits for eligibility and the ceilings on allowable deductions are higher in Alaska and Hawaii than in the other states due to their higher cost of living.

The Census Bureau obtains monthly totals of food stamp recipients for states from the U.S. Department of Agriculture. After releasing the revised 1993 SAIPE state and county estimates but before preparing the 1995 estimates, the Census Bureau conducted research to determine how best to use these data for input to the state regression model. Based on that research, the Bureau decided to use the monthly counts averaged over a 12-month period centered on January 1 of the calendar year subsequent to the income reference year to form the predictor variable in the model. (Previously the data used were for July of the reference year.) The Census Bureau further refined the food stamp counts in three ways: it subtracted counts by state of the numbers of people who received food stamps due to specific natural disasters from the counts of the total number of recipients; it used the results of time-series analysis of monthly food stamp data from October 1979 through September 1997 to smooth outliers; and it adjusted the counts of food stamp recipients in Alaska and Hawaii downward to reflect the higher eligibility thresholds for those states.

For counties, the Census Bureau obtains counts of food stamp recipients from USDA and, in some instances, from state agencies. However, the information obtained for each county is not always the same: in most counties, the counts of food stamp recipients pertain to July; for some counties, they are an average of the monthly counts for the year. In developing the 1995 county regression model, the Census Bureau raked the county food stamp numbers to the adjusted state food stamp numbers.

---

[13]The poverty guidelines used for determining program eligibility are derived by smoothing the official poverty thresholds for families of different sizes (see Fisher, 1992).

Although the Food Stamp Program is generally administered uniformly across all states, estimated participation rates–the proportion of eligible households that apply for and receive benefits–are not the same. Differences in participation rates, which may stem from differences in outreach efforts, the stigma associated with participation, or other factors, could affect the comparability of food stamp counts across areas in terms of how well they relate to poverty.

Income Tax Returns

The Census Bureau uses information from federal income tax returns to construct family units and determine the number of child exemptions in families that report incomes below the applicable poverty thresholds on their returns. Individual tax returns are assigned to counties on the basis of their address information. If the address is a post office box or rural route number and not a city-style address, such as 104 Main St., the ZIP code is used to assign the address to a county. There are three major advantages of data from tax returns: (1) coverage of a very large proportion of the population, (2) coverage of a very large proportion of the income received by families, and (3) the availability of data on an annual basis.

After releasing the revised 1993 state and county estimates and before developing the 1995 estimates, the Census Bureau discovered and corrected an error in processing the 1989 IRS data. The corrected data were used to reestimate the decennial census equation that provides the residual predictor variable in the 1995 state model (see Chapter 4). The corrected data were also used to reestimate the 1989 state and county models for evaluation purposes. In both the state and county models, child exemptions reported by families on tax returns were redefined to include children away from home in addition to children at home. This change may increase the number of IRS poor child exemptions in households with children away from home both because of the additional children and because poverty thresholds are higher for larger size families.

The number of child exemptions reported on tax returns for families with incomes below the poverty threshold, like the number of food stamp recipients, is an imperfect measure of poverty for school-age children. Not all people file tax returns, especially those with very low incomes or income mostly from nontaxable sources. In addition, "income" as defined on tax returns does not include all the sources of income that are used in the official measure of poverty, and tax filing units are not totally consistent with the Census Bureau's definition of families. Moreover, the address on a tax return does not always correspond to a filer's residential address. Also, from evaluation, the Census Bureau has found some differences between states in the completeness of the tax files that it obtains from IRS that may affect use of the data in models (Cardiff, 1998). These differences occur because the Census Bureau receives an early version of the data for each tax filing year from the IRS. Nonetheless, tax information, like counts of food stamp recipients, is a useful variable to develop predictions of poverty for school-age children.

TIMELINESS OF ESTIMATES

The CPS provides more timely data than the decennial census; however, SAIPE estimates of poor school-age children for counties that are derived from the CPS will not be current. Thus, the Census Bureau released SAIPE county estimates of poor school-age children in early 1997 for income year 1993 and county and school district estimates in early 1999 for income year 1995, for use for Title I allocations for the next two school years (1999-2000 and 2000-2001 in the case of the 1995 school district estimates).[14]

The reason for the lag between the income reference year and the year of release of estimates is that most data used in the county model are not available until 2 years after the period to which they refer. The time lag is also caused by the decision to use 3 years of CPS data in the Census Bureau's model to improve the precision of the estimates. The lag means that the estimates will not capture any changes in the extent and distribution of poverty among school-age children that may have occurred since the year to which they apply.

Published CPS data indicate that poverty among school-age children for the nation as a whole increased from 17.4 percent in 1989 to 20.1 percent in 1993 and then declined to 15.5 percent in 1999 (U.S. Census Bureau, 1990:Table 18; 1995a:Table 8; 2000:Table 2).[15] Data are not available for 2000, and no data are readily available with which to estimate the changes in the distribution of poverty among school-age children across states and counties that may have occurred since the last release of estimates.

The panel was asked to evaluate the accuracy of the updated county-level estimates that the Census Bureau was able to produce with available data. The panel addressed the question of the accuracy of the estimates for the estimation year (1993, 1995), not the question of how well the estimates for 1993 (1995) predict poverty among school-age children in 1997 (1999). It should be a priority for research and development by the Census Bureau to determine ways to reduce the lag between the time period of the estimates and the year of their release (see Chapter 9).

---

[14]Estimates for income year 1997 are scheduled for release in fall 2000.

[15]These estimates are for related children aged 6-17; estimates are not published for related children aged 5-17.

4

Estimation Procedure for Counties

The task for the SAIPE Program of producing reasonably reliable and current county-level estimates of poor school-age children for Title I allocations is a challenging one. At present, no single administrative records or survey data source provides sufficient information with which to develop reliable direct county estimates of the numbers and proportions of poor school-age children that are more up to date than census estimates. The March Income Supplement to the Current Population Survey (CPS) can provide reasonably reliable annual direct estimates of such population characteristics as the number and proportion of poor children at the national level and possibly for the largest states. However, the CPS cannot provide direct estimates for the majority of counties because the sample does not include any households in them. And for almost all of the counties with households in the CPS sample (1,274 of a total of 3,142 counties in 1995), the estimates have a high degree of sampling variability.[1] Nonetheless, the CPS data may serve as the basis for creating usable estimates for counties through the application of statistical estimation techniques to develop "model-based" or "indirect" estimates.

Model-based or indirect estimators use data from several areas, time periods, or data sources to "borrow strength" and improve the precision of estimates for

[1]For a description of the March CPS and differences between income and poverty data from the CPS and the 1990 census long-form sample, see Chapter 3. The 1990 census sample includes households in all counties and covers 15.7 million households, 30 times more than the 50,000 households in the CPS; even the 1990 census estimates are highly variable for small counties (Table 3-1).

small areas. A model-based approach is needed when there is no single data source for the area and time period in question that can provide direct estimates of sufficient reliability for the intended purpose. The Census Bureau has used this strategy to develop estimates of median family income for states (Fay, Nelson, and Litow, 1993). In the 1970s, it used model-based methods to improve 1970 census small-area income estimates for use in developing updated per capita income estimates for governmental jurisdictions (Fay and Herriot, 1979) and, in part, to develop population estimates for states and counties (see Spencer and Lee, 1980).

This chapter provides a summary description and evaluation of the model-based approach used by the Census Bureau to develop estimates by county of the numbers and proportions of school-age children in families in 1996 who were poor in 1995 (referred to as the 1995 county estimates). The estimation procedure involves the use of separate county and state regression models.[2] The chapter also summarizes differences between the state and county models used to develop the 1995 county estimates and the models used to develop the original and revised 1993 county estimates. Additional detailed documentation for the 1993 state and county models (and alternative models) is provided in Appendix A; see also Bell et al. (2000). For the county model, see also Coder, Fisher, and Siegel (1996) and Fisher (1997); for the state model, see also Fay (1996) and Fay and Train (1997). The Census Bureau's web site (www.census.gov/hhes/www/saipe.html) provides an overview of the estimation procedures and contains a number of papers on the SAIPE methods.

When the Department of Education uses the Census Bureau's school district estimates of poor school-age children for direct allocation of Title I funds to districts, county estimates are not used directly in the allocation process. However, the county estimates are critical to the development of school district estimates. As a result of the lack of data at the school-district level, the Census Bureau is constrained to use for school districts a very simple model-based method referred to as a shares method, which, for 1995 estimates, applied the shares or proportions of poor school-age children for the school districts in a county according to the 1990 census to the updated 1995 county estimates to obtain updated school district estimates (see Chapter 7). Therefore, in order to evaluate the 1995 school district estimates, it is essential to understand and evaluate the 1995 county estimates.

[2]The panel's final report provides a more mathematical presentation of the development of these models (National Research Council, 2000:Ch.3).

1995 ESTIMATION PROCEDURE

The Census Bureau's 1995 estimation procedure for counties includes the following steps:

(1) Developing and applying the Census Bureau's county model to produce initial estimates of the numbers of poor school-age children. The county estimation process involves:

- obtaining data from administrative records and other sources that are available for all counties to use as predictor variables;
- specifying and estimating a regression equation that relates the predictor variables to a dependent variable, which is the estimated log number of poor school-age children from 3 years of the March CPS for counties with households with poor school-age children in the CPS sample; and
- using the estimated regression coefficients from the equation and the predictor variables to develop estimates of poor school-age children for all counties. For counties with households in the CPS sample, the predictions from the model are then combined by a "shrinkage" procedure with the CPS estimates for those counties.

(2) Developing and applying the Census Bureau's state model to produce estimates of the numbers of poor school-age children by state. The state estimation process is similar to that for counties, although the state model differs from the county model in several respects.

(3) Adjusting the initial estimates of poor school-age children from the county model (step 1) for consistency by state with the estimates from the state model (step 2) to produce final estimates of the numbers of related children aged 5-17 in poverty by county.

In addition, the Census Bureau produces various state and county population estimates, which are used in the estimation of poor school-age children (see Chapter 8).[3] Finally, the Census Bureau produces separate estimates of poor school-age children for Puerto Rico, which is treated as a county and school district equivalent in the Title I allocation formulas (see Appendix E).

[3]The county population estimates of the total number of school-age children were used by the Department of Education to calculated estimated proportions of poor school-age children when it made Title I allocations to counties in the two-stage process that was used through the 1998-1999 school year (see Chapter 2).

Step 1: 1995 County Model

County Equation

The 1995 county equation uses as predictor variables county estimates from Internal Revenue Service (IRS) records for 1995, Food Stamp Program records for 1995,[4] the 1990 census, and the Census Bureau's postcensal population estimates program for 1996. As the dependent or outcome variable, it uses county estimates of the number of poor school-age children averaged over 3 years of the March CPS (data from the March 1995, 1996, and 1997 CPS, covering income in 1994, 1995, and 1996). The equation takes the following form:

$$z_i = \beta_0 + \beta_1 w_{1i} + \beta_2 w_{2i} + \beta_3 w_{3i} + \beta_4 w_{4i} + \beta_5 w_{5i} + v_i + a_i \ , \qquad (1)$$

where:

z_i = log(3-year weighted average of number of poor school-age children in county i based on 3 years of March CPS data),
w_{1i} = log(number of child exemptions reported by families in poverty on tax returns in county i),
w_{2i} = log(number of people receiving food stamps in county i),
w_{3i} = log(estimated population under age 18 in county i),
w_{4i} = log(number of child exemptions on tax returns in county i),
w_{5i} = log(number of poor school-age children in county i in the previous census),
v_i = model error for county i, and
a_i = sampling error of the dependent variable for county i.

Dependent Variable The Census Bureau originally decided to model the *number* of poor school-age children, instead of the *proportion*, because of concern that the county population estimates of school-age children that would form the basis for converting the estimated proportions to estimated numbers were of uncertain quality. Hence, it would be difficult to construct estimates of the precision of the estimated numbers of poor school-age children at the county level, which played the most important role in the Title I allocation formula under the two-stage procedure.

The Census Bureau decided to estimate the number of poor school-age children *at a particular time* and not to estimate the *change* in the number since the 1990 census because it concluded that the available administrative data were

[4]The food stamp data for most counties pertain to July 1995; for other counties, they are an annual average of monthly counts for 1995. The county numbers are controlled to state food stamp estimates, which are 12-month averages centered on January 1996 (see Chapter 3).

likely to be measured more consistently across areas at a given time than they would be over time, given changes in tax and transfer programs.

The Census Bureau decided to combine 3 years of CPS data to form the dependent variable for the county model. The combination of years improves the precision of the dependent variable, although the dependent variable consequently pertains to the 3-year period rather than to the estimation year.

The weighted 3-year average of the number of poor school-age children in each county is computed as the product of the weighted 3-year average estimated CPS poverty rate for related children aged 5-17 and the weighted 3-year average estimated CPS number of related children aged 5-17 for that county. The weights for these averages are the fractions of the 3-year estimated total of CPS interviewed housing units in each county that contain children aged 5-17 in each year.

Because only a subset of counties have households in the March CPS sample, the relationships between the predictor variables and the dependent variable in the model are estimated solely from this subset of counties. This subset includes proportionately more large counties and proportionately fewer small counties than the distribution of all counties. Also, because the dependent variable is measured on a logarithmic scale for reasons given below and values of 0 cannot be transformed into logarithms, a number of counties whose sampled households contain no poor school-age children are excluded from the estimation. In all, 985 of 3,142 counties were included in the 1995 model estimation–the remainder were excluded because none of their CPS-sampled households had school-age children who were poor (262 counties), none of their CPS-sampled households had school-age children (27 counties), or they had no CPS-sampled households (1,868 counties). Corresponding figures for 1993 are as follows: 1,184 of 3,143 counties were included in the model estimation; 304 counties had CPS households with school-age children but none with school-age children who were poor, 41 counties had CPS-sampled households but none with school-age children, and 1,614 counties had no CPS-sampled households at all.[5]

Predictor Variables The choice of predictor variables was governed by data availability and the assumed relationship of the variables to poverty. The number of child exemptions reported by families in poverty on tax returns and the number of food stamp recipients were included as variables that are indicative of poverty and available on a consistent basis (or reasonably consistent basis, in the case of

[5]The reason why the 1993 model estimation included a higher proportion of counties than the 1995 model estimation is because a redesign of the CPS sample was phased in between April 1994 and July 1995. Some counties were included in the old design but not the new and vice versa; the estimation included all counties that had at least 1 year of CPS data in the 3 years centered on the estimation year.

food stamps) for all counties in the nation.[6] The 1990 census estimate of poor school-age children was used in the 1995 model on the assumption that previous poverty is likely to be indicative of subsequent poverty. The total number of child exemptions on tax returns and the population estimate of the total number of children under 18 were included in order to cover children not reported on tax returns (i.e., in nonfiling families), who are assumed to be poorer on average than other children. (The estimated regression coefficients for the 1995 county model predictor variables are given in Table 6-2.)

Form of the Variables The dependent variable and all of the predictor variables are measured on a logarithmic scale. A reason to use logarithms is the wide variation in the CPS estimates of the dependent variable and the values of the predictor variables among counties when they are measured on the numeric scale: transforming the variables to logarithms made their distributions more symmetric and the relationships between some of them and the dependent variable more linear.

Estimation of Model and Sampling Error Variance The total squared error of the county estimates (the difference between the model estimates and the direct estimates from the CPS) has two sources: model error (v) and sampling error (a), which are the last two terms in the county equation.[7] Model error is the difference between the value of the logarithm of the 3-year weighted average of the number of poor school-age children that would have been obtained had all the households in the county been included in the CPS sample and the model estimate of this quantity. Sampling error is the difference between the estimate of this quantity from the CPS sample and the value that would have been obtained had all households in the county been included in the CPS sample. Model error is assumed to be constant across counties (see below). Sampling error is not constant across counties; it is larger for counties that have fewer households included in the CPS sample.

Because a procedure to estimate the sampling error variance directly for the March CPS has not yet been developed (see Chapter 9), the variances of the model error and sampling error terms in the 1995 county equation are estimated in a multiple-step process that involves several assumptions. First, equation (1) is estimated for 1989, using the 1990 census estimate of poor school-age children as the dependent variable and 1989 IRS and food stamp data, 1990 census popula-

[6]Poverty status for families on tax returns is determined by comparing the adjusted gross income on each return to the average poverty threshold for the total number of exemptions on the form. Although there are differences between the CPS and IRS definitions of income and family composition, they are not critical for purposes of developing a predictive model.

[7]As used in statistics, "error" is the inevitable discrepancy between the truth and an estimate due to variability in measurements and the fact that model predictions are imperfect.

tion data, and 1980 census poverty data as the predictor variables. The estimation procedure includes almost all counties, excluding only those few that had no poor school-age children in the long-form sample in 1980 or 1990 or that did not exist in both 1980 and 1990. A generalized variance function is used to estimate the sampling variances of the census estimates, which are often relatively small because of the large size of the census long-form sample. Then, by estimating equation (1) using weighted least squares in an iterative process, in which a starting value is specified for the model error variance and the sampling error variances are known, maximum likelihood estimates of the regression coefficients and the model error variance are obtained. Most of the mean square error in the census equation (about 90%) is derived from model error variance.

It is assumed that the model error variance for the CPS equation for 1995 is the same as that for the 1990 census equation and that it has the same value for each county. Then, the CPS equation is estimated by iteratively weighted least squares, which produces maximum likelihood estimates of the regression coefficients and the total sampling error variance, which is distributed among the counties as an inverse function of their sample size.[8] Most of the CPS mean square error (about 90%) is derived from sampling error variances.

The resulting estimates of model error variance and sampling error variance are used to determine the weights to give to the model prediction from the maximum likelihood procedure and to the CPS direct estimate in developing estimates of poor school-age children for counties with sampled households with poor school-age children in the CPS.

Combining the County Equation and CPS Estimates

By calculating the relationships among the predictor variables and the CPS estimates of school-age children in poverty for the subset of counties that have households with poor school-age children in the March CPS sample, it is possible to obtain a good estimate of a regression equation for predicting the number of poor school-age children in a county, even though the CPS estimates for many small counties have large levels of uncertainty. The regression equation can then be used to predict the number of school-age children in poverty from the food stamp, IRS, population estimates, and previous census predictor variables for each county, whether or not the county is in the March CPS sample.

For counties that have households with poor school-age children in the March CPS sample, a weighted average of the model prediction and the estimate based on data from the sampled households (the direct estimate) is used to produce an estimate for that county using empirical Bayes (“shrinkage”) procedures for com-

[8]The weights used are the reciprocal of the sum of the estimated sampling variance of the estimate of the log number of poor school-age children in a given county plus the estimated model error variance, assumed to be constant across counties; see Appendix A.

bining estimates (see Fay and Herriot, 1979; Ghosh and Rao, 1994; Platek et al., 1987; Rao, 1999). The weights that are given to the model prediction and the direct estimate depend on their relative precision (see discussion above of how model error variance and sampling error variance are estimated). For a county with very few sample households in the CPS and hence a high level of sampling variability in the direct estimate, most of the weight will be given to the model prediction and little to the direct estimate. For a county with a larger number of sampled households in the CPS, more weight will be given to the direct estimate and less to the model prediction. For almost all counties that have households with poor school-age children in the CPS, most of the weight is given to the model prediction; for the 1993 estimates the weight for the model prediction was less than 0.5 for only 2 counties; it was less than 0.75 for only 13 counties. For counties that lack households with poor school-age children in the CPS sample, the prediction from the model is the estimate. After shrinkage, the initial county estimates are obtained by transforming the shrunk values from the logarithmic to the numeric scale.

Step 2: 1995 State Model

State Equation

The state model equation takes the following form:

$$y_j = \alpha_0 + \alpha_1 x_{1j} + \alpha_2 x_{2j} + \alpha_3 x_{3j} + \alpha_4 x_{4j} + u_j + e_j \ , \qquad (2)$$

where:

y_j = estimated proportion of school-age children in state j who are in poverty based on the March CPS that collects income data pertaining to the estimation year,[9]

x_{1j} = proportion of child exemptions reported by families in poverty on tax returns in state j,

x_{2j} = proportion of people receiving food stamps in state j,

x_{3j} = proportion of people under age 65 not included on an income tax return in state j,[10]

[9]The numerator is the estimated number of poor related children aged 5-17 from the CPS; the denominator is the estimated total population of children aged 5-17, whether or not they are related to a family, from the CPS. (See text for the reason to include unrelated children in the denominator; that denominator, however, excludes the institutionalized, who are not sampled.)

[10]This percentage is obtained by subtracting the estimated number of exemptions on income tax returns for people under age 65 from the estimated total population under age 65 derived from the Census Bureau's population estimates program; see Chapter 8.

x_{4j} = residual for state j from a regression of the proportion of poor school-age children from the prior decennial census on the other three predictor variables (x_{1j}, x_{2j}, x_{3j}),
u_j = model error for state j; and
e_j = sampling error of the dependent variable for state j.

Differences from the County Equation

The Census Bureau's state model for estimates of poverty among school-age children is similar to the county model. However, it differs in a number of respects:

Dependent Variable The state model uses the proportion of school-age children in poverty in each state as the dependent variable: that is, the dependent variable is a poverty ratio rather than the number of poor school-age children, as in the county model.[11] The numerator for the ratio is the CPS estimate of poor school-age children in a state (i.e., the estimate of the number of poor related children aged 5-17); the denominator is the CPS estimate of the total number of children aged 5-17 in the state. A different denominator–total CPS school-age children, rather than the slightly smaller universe of related school-age children–is used for consistency with the population estimates that are available to convert the estimated poverty ratios to estimated numbers of poor school-age children.

In addition, the dependent variable in the state model is derived from 1 year of CPS data (the March 1996 CPS for the 1995 model), rather than a 3-year average as in the county model. This decision was made because the sample sizes for states are all reasonably large for the purpose of fitting the regression model.

Predictor Variables As can be seen above, the state model uses a somewhat different set of predictor variables than the county model. (The estimated regression coefficients for the state model predictor variables are given in Table 6-7.) The state model includes a predictor variable that is the residual from a regression of the proportion of poor school-age children from the prior decennial census on the other three predictor variables. During the development of the state model, the Census Bureau determined that there was a correlation between the residuals from estimating the model for 1979 with 1980 census data and the residuals from estimating the model for 1989 with 1990 census data. In other words, states that had more poverty than predicted by the cross-sectional model for 1979 also tended to have more poverty than predicted by the cross-sectional model for

[11]The dependent variable is termed a ratio because the denominator is not exactly the same as that for the official published poverty rates.

1989. This result was used to improve the model predictions by including the residual from a regression for the prior census as one of the predictor variables.

Form of the Variables The variables in the state model are proportions rather than numbers and are not transformed to a logarithmic scale as is done in the county model.[12] A log-based model was examined, but the Census Bureau decided not to transform the variables because, unlike the situation with the county model, the state-level distributions of the estimated proportions for the predictor variables are reasonably symmetric, and the relationships of the state-level estimated proportions with the dependent variable are approximately linear.

Combining the State Equation and CPS Estimates

All states have sampled households in the CPS; however, the variability associated with estimates from the CPS is large for some states. As is done for the initial county estimates, the predictions from the state model and the CPS estimates are weighted according to their relative precision to produce estimates of the proportion of poor school-age children in each state. To produce estimates of the number of poor school-age children in each state, the estimates of the proportion poor are multiplied by estimates of the total number of noninstitutionalized school-age children from the Census Bureau's population estimates program. (The estimates of noninstitutionalized school-age children, which include some adjustments for residents of military group quarters and college dormitories, are the closest approximation available to the CPS estimates of school-age children.) Finally, the state estimates of the number of poor school-age children are adjusted to sum to the CPS national estimate of related school-age children in poverty. This adjustment is a minor one; for 1995 it changed the state estimates by less than 0.5 percent; for 1993, the adjustment changed the state estimates by less than 1 percent.

Step 3: Combining the County and State Estimates

The final step in developing estimates of numbers of poor school-age children by county is to adjust the initial estimates from the county model (after shrinkage and transformation to the numeric scale with a correction for transformation bias,[13] step 1) for consistency with the estimates from the state model

[12]The estimates that are transformed into logarithms in the county model are numbers, not proportions. However, evaluation determined that, if the county model were to estimate proportions, a logarithmic transformation of the dependent and predictor variables would be helpful in that case as well (see Chapter 5).

[13]Transformation bias occurs when a regression model estimates an expected value for the depen-

(after shrinkage, step 2). The estimate for each state from the state model is then divided by the sum of the estimates for each county in that state to form a state raking factor. Each of the county estimates in a state is multiplied by the state raking factor so that the sum of the adjusted county estimates equals the state estimate. For the final county estimates of poor school-age children in 1995, the average state raking factor was 0.97; two-thirds of the factors were between 0.88 and 1.06. For the final, revised county estimates of poor school-age children in 1993, the average state raking factor was 1.065; two-thirds of the factors were between 0.975 and 1.154.

DIFFERENCES BETWEEN 1995 AND 1993 ESTIMATION PROCEDURES

The procedure summarized above to produce the 1995 county estimates that the Census Bureau released in early 1999 differs in a few respects from the procedure that was used to produce the revised 1993 estimates that the Bureau released in early 1998. The changes involved the input data for the state and county models:

- An error in processing the 1989 IRS data was discovered and corrected. The corrected data were used to reestimate the decennial census equation that provides the residual predictor variable in the 1995 state model (x_{4j} in equation (2)). The corrected data were also used to reestimate the 1989 state and county models for evaluation purposes (see Chapter 6).
- Several changes were made to the food stamp data for input to the state model: instead of using data for July of the estimation year, the number of food stamp recipients was changed to a 12-month average centered on January 1 of the following year; counts by state of the numbers of people who received food stamps due to specific natural disasters were obtained from the Department of Agriculture and subtracted from the counts of the total number of recipients; time-series analysis of monthly state food stamp data from October 1979 through September 1997 was used to smooth outliers; and food stamp recipient data for Alaska and Hawaii were adjusted downward to reflect the higher eligibility thresholds for those states.

dent variable that is on a different scale from that for which estimates are needed. In this instance, the county model predicts poor school-age children on the log scale; when the predictions on the log scale are exponentiated back to the original numeric scale, the result is the exponential of the expected value of the dependent variable on the log scale, which is different from the expected value of the dependent variable on the original scale. This difference is referred to as transformation bias, for which a correction is made.

- The food stamp numbers for the county model were raked to the adjusted state food stamp numbers.
- In both the state and county models, child exemptions reported by families on tax returns were redefined to include children away from home in addition to children at home. This change may increase the number of IRS poor child exemptions in households with children away from home both because of the additional children and because poverty thresholds are higher for larger size families.

DIFFERENCES BETWEEN ORIGINAL AND REVISED 1993 ESTIMATION PROCEDURES

The procedure used to produce the revised 1993 county estimates that the Census Bureau released in early 1998 differed in some respects from the procedure used to produce the original 1993 estimates that the Bureau released in early 1997. The principal difference involved a change in one of the predictor variables in the county regression model.

The changes listed below in producing the revised 1993 estimates were retained in producing the 1995 estimates. Specifically:

- The revised county model includes the population under 18 as a predictor variable; the original county model included the population under 21 as a predictor variable. The purpose of this variable (whether for the population under 18 or under 21) is to estimate–in conjunction with the variable measuring total child exemptions on IRS tax returns–the number of children in families that did not file a tax return. Evaluation determined that the original estimation procedure was not working well for counties with large numbers of people under age 21 in group quarters, primarily college students and military personnel. Specifically, the model was overpredicting the number of poor school-age children for those counties. Limiting the predictor variable to the population under 18 reduced the bias in the model predictions for counties classified by percent group quarters residents and improved the model predictions in other respects (see Chapter 6).
- Examination of the pattern of residuals (differences between the model predictions and the direct estimates) for counties with sampled households in the March CPS indicated that the original method for estimating model error variance and sampling error variance (described above) was not working as well as it should. The variability of the standardized residuals increased with the number of CPS sample cases rather than remaining constant, and this pattern was common to a variety of alternative models that were examined. The revised 1993 county model includes a slight revision to the procedure for estimating the sampling error variance, which moderated but did not eliminate the anomalous pattern. Further work will be required to further reduce the problem (see Chapter 9).

However, improving the estimation of the model error and sampling error variances will probably have only a limited effect on the county estimates. The main use of these variance estimates is to determine the weights to be given to the model predictions and to the CPS direct estimates in forming estimates for counties that have sampled households with poor school-age children in the CPS. Since the model predictions are the dominant component of the county estimates in most cases, changing the weights will not have a substantial impact.

- The original model was estimated using a method-of-moments procedure; for the revised model, it was decided that maximum likelihood estimation would be used. This change had only a small effect on the estimated regression coefficients for the predictor variables. The main effect of the change was to increase the estimated sampling error variance. Hence, in comparison with the original 1993 estimates, the revised 1993 model predictions are given somewhat more weight and the CPS direct estimates are given somewhat less weight when weighted estimates are formed for counties that have sampled households with poor school-age children in the CPS. However, this difference had relatively little effect on the county estimates.

5

Alternative County Models

The Census Bureau's procedure for developing updated county estimates of poor school-age children uses a county model, a separate state model, and county population estimates. All three components are important, but the heart of the estimation procedure is the county model. The task of developing good county estimates of poor school-age children is more difficult than the task of developing good state estimates of poor school-age children or good county estimates of total school-age children. Hence, the evaluation efforts of the panel and the Census Bureau focused mainly on the county model.

In selecting a specific model for developing small-area poverty estimates that are to be used for such an important public purpose as allocating funds, it is important to compare the selected model to competing models that may have specific advantages. When the original county estimates of poor school-age children in 1993 were released in early 1997, the Census Bureau had not had time to undertake a thorough assessment of the performance of the model used or to compare its performance to that of other models. Subsequently, the panel and the Census Bureau developed a range of alternative county models to evaluate. In a first round of evaluations, 12 models were examined. On the basis of the results of those evaluations, a second round of evaluations examined four models that appeared practicable to use in the SAIPE Program in the near term.

The basic features of the alternative models that were examined are summarized below. All of the models were estimated for 1993, and all except the bivariate models were estimated for 1989 to provide estimates for external evaluation by comparison with 1990 census estimates.[1]

MODEL CHARACTERISTICS

The alternative county models that the Census Bureau and the panel examined are distinguished broadly by three characteristics: (1) treatment of information from the previous census–whether the model includes a predictor variable from the previous census in a single equation or uses a bivariate formulation that links a census equation with a CPS equation; (2) the form of the variables–whether they are numbers or proportions, transformed to logarithms or untransformed; and (3) whether the model includes intercept terms for each state (i.e., fixed state effects).

Treatment of Information from the Previous Census The revised county model that the Census Bureau used to produce estimates of poor school-age children in 1993 and 1995 is a single-equation model in which the dependent variable is from the CPS and one of the predictor variables is the estimated number of poor school-age children from the previous census. The inclusion of the census predictor variable is based on the assumption that poverty in a prior year is indicative of poverty in a later year.

The state model makes use of information from the previous census in a different way. The state model equation, in which the dependent variable is also from the CPS, includes a predictor variable that is the estimated residual from a similar regression for the previous census. The underlying assumption is that states that had more (less) poverty than predicted for the census year will continue to have more (less) poverty for a later year than the model would predict without the residual variable. This assumption was supported by an analysis that showed the residuals from a state model estimated for 1979 with 1980 census data to be correlated with the residuals from a state model estimated for 1989 with 1990 census data.

The possible advantage of having the county model include the estimated residual from an equation for the previous census could not be established because the necessary administrative data are not available with which to estimate a county equation from the 1980 census (for 1979). As an alternative, the Census Bureau developed a bivariate formulation of the county model in order to make

[1]For technical information on the models included in the first round of evaluations, see Appendix A. The models specified do not exhaust the list of possibilities, but they are a reasonable range of alternatives to consider at the present time. See Chapter 9 and National Research Council (2000:Ch.3) for model formulations that could be considered as part of a longer term research program for producing small-area poverty estimates.

more complete use of information from the previous census in a manner analogous to the state model (Bell, 1997a). In the bivariate formulation for 1993, the county model jointly estimates two regression equations, one that produces 1993 county estimates based on March 1993-1995 CPS data and the other that produces 1989 county estimates based on 1990 census data. This formulation incorporates information from the census by allowing the model errors in the two equations to be correlated (see below, "Bivariate Models").

Form of the Variables In the revised county model, the dependent variable is the log number of poor school-age children, and the predictor variables are also numbers that are transformed to logarithms. The Census Bureau and the panel examined alternative county models in which the dependent variable is the proportion, or rate, of poor school-age children. For some of these rate models, the dependent and predictor variables are transformed to logarithms; for others, they are not transformed. Models for which the dependent and predictor variables are untransformed numbers were not considered because, when not transformed to logarithms, the distributions of the dependent and predictor variables at the county level have a wide range and are not symmetric; also, the predictor variables do not have linear relationships with the dependent variable. Untransformed poverty rates do not share these problems to the same extent, although it is possible to obtain predicted negative values from an untransformed formulation.

Inclusion of Fixed State Effects In the revised county model, there are no predictor variables that explicitly account for regional or state effects. After the county estimates are produced from the model, combined with the direct CPS estimates where applicable, and transformed to the numeric scale, they are raked for consistency with the estimates from the state model. Analysis of the size and variability of the raking factors (see Chapter 6) suggested that the county model may not adequately account for differences among states in the relationship of the predictor variables to the dependent variable and, consequently, that the county model may not adequately reflect the variation among counties within a state.

As a way to explore this problem, the Census Bureau developed a fixed state effects model by including an indicator, or dummy, variable for each state. The purpose of these state indicator variables is to enable the model to more accurately capture the variation among counties within each state by accounting for differences in the level of the dependent variable by state.

MODELS EXAMINED IN THE FIRST ROUND OF EVALUATIONS

Of the 12 models examined in the first round of evaluations, 6 were single-equation models, and 6 were bivariate models. Nine of the 12 models transform the values of the dependent and predictor variables into logarithms. Because logarithms cannot be taken for values of 0, these models are estimated only for the counties with sampled households in the CPS that contain at least one poor school-age child: 1,184 of 3,143 counties for the 1993 models. The other three models, which do not transform the variables (all three are rate models), use data for all counties with sampled households in the CPS that contain at least one school-age child: 1,488 counties for the 1993 models. A topic for future work is how to use all counties with CPS-sampled households with school-age children in estimating a log-based model (see Chapter 9).

Single-Equation Models

The basic form of a single-equation county model is

$$z_i = \beta_0 + \beta_1 w_{1i} + \beta_2 w_{2i} \ldots + \beta_5 w_{5i} + v_i + a_i \ , \tag{1}$$

where:

z_i = the dependent variable in county i (log number or proportion of poor school-age children from 3 years of CPS data),
$w_{1i} \ldots w_{5i}$ = the predictor variables in county i,
v_i = model error for county i, and
a_i = sampling error of the dependent variable for county i.

The formulation with fixed state effects adds a set of indicator variables, one for each state. The indicator for a given state is 1 for all counties in that state and 0 otherwise. The intercept term, β_0, is dropped from the models with fixed state effects to avoid overidentification. The addition of a large number of dummy variables does not result in too few degrees of freedom because more than 1,000 counties are used to fit the regression coefficients.

Six single-equation models were evaluated in the first round (see Table 5-1):

(1) ***Log Number Model (Under 21)*** The dependent variable is the CPS estimate of the log number of poor school-age children, derived by multiplying for each county the 3-year weighted average poverty rate for related children aged 5-17 by the 3-year weighted average of total related children aged 5-17. The predictor variables, all of which are transformed to logarithms, are the number of child exemptions (assumed to be under age 21) reported by families in poverty on tax returns; the number of people receiving food stamps; the estimated population under age 21; the total number of child exemptions on tax returns; and the estimated number of poor school-age children in the 1990 census. For the 1993 model, the IRS and food stamp data pertain to 1993; the population estimates data pertain to 1994. This is the original model used by the Census Bureau to produce 1993 county estimates of poor school-age children.

TABLE 5-1 Single-Equation County Models: Dependent Variable and Predictor Variables

Model	Dependent Variable, z_i	Predictor Variables, $w_{1i} \ldots w_{5i}$	Form of the Predictor Variables
(1) Log Number (Under 21)	Log 3-year weighted average number of poor school-age children	(1) Number of child exemptions reported by families in poverty on tax returns (2) Number of people receiving food stamps (3) Population under 21 (4) Number of child exemptions on tax returns (5) Number of poor school-age children in 1990 census	Transformed to logarithms
(2) Log Number (Under 18)	Log 3-year weighted average number of poor school-age children	(1) Number of child exemptions reported by families in poverty on tax returns (2) Number of people receiving food stamps (3) Population under 18 (4) Number of child exemptions on tax returns (5) Number of poor school-age children in 1990 census	Transformed to logarithms
(3) Log Number (Under 21) with Fixed State Effects	Same as Log Number Under 21 (1)	Same as Log Number Under 21 (1) with the addition of state indicator variables	Transformed to logarithms

(4) Log Rate	Log poverty ratio for school-age children (3-year sum of poor related children aged 5-17 divided by 3-year sum of total CPS children aged 5-17)	(1) Ratio of number of child exemptions reported by families in poverty on tax returns to total number of child exemptions on tax returns (2) Ratio of number of people receiving food stamps to total population (3) Ratio of total number of child exemptions on tax returns to total population under 21 (4) Ratio of number of poor related children aged 5-17 to total number of related children aged 5-17 from the 1990 census	Transformed to logarithms
(5) Rate	Poverty ratio for school-age children (same as Log Rate (4), except untransformed)	Same as Log Rate (4), except untransformed	Untransformed
(6) Hybrid Rate-Number	Same as Log Rate (4)	Same as Log Number Under 21 (1)	Transformed to logarithms

NOTE: The models are estimated for 1993 from 3 years of CPS data (March 1993, 1994, and 1995, covering income in 1992, 1993, and 1994).

(2) ***Log Number Model (Under 18)*** The dependent and predictor variables are the same as in (1), except that the estimated population under age 18 replaces the estimated population under age 21. This is the revised model used by the Census Bureau to produce the revised 1993 and the 1995 county estimates of poor school-age children (see Chapter 4). It was included in the first round of evaluations after it became apparent that the log number model (under 21) was not performing well for counties with large numbers of people under age 21 in group quarters (see Chapter 6).

(3) ***Log Number Model with Fixed State Effects*** The dependent and predictor variables are the same as in (1), with the addition of state indicator variables.

(4) ***Log Rate Model (Under 21)*** The dependent variable is the CPS estimate of the log proportion poor, or log poverty rate, for school-age children: more precisely, a poverty ratio–similar to the state model–in which for each county the numerator is the sum over 3 years of the estimated number of poor related children aged 5-17 and the denominator is the sum over 3 years of the estimated total number of CPS children aged 5-17. The predictor variables are also ratios: the ratio of the number of child exemptions reported by families in poverty on tax returns to the total number of child exemptions on tax returns; the ratio of the number of people receiving food stamps to the total population (all ages); the ratio of the total number of child exemptions on tax returns to the total population under age 21;[2] and the ratio of the estimated number of poor related children aged 5-17 to the estimated total number of related children aged 5-17 from the 1990 census. All variables are transformed to logarithms.

(5) ***Rate Model*** The dependent variable and predictor variables are the same as in (4), but all variables are ratios, untransformed.

(6) ***Hybrid Log Rate-Number Model*** The dependent variable is the CPS estimate of the poverty ratio for poor school-age children as in (4); the predictor variables are the same as in (1); that is, they represent numbers, not ratios; and all variables are transformed to logarithms.

---

[2]In 292 counties, the ratio of total child exemptions on tax returns to the total noninstitutionalized population under age 21 in 1993–the tax filer population ratio–is greater than 1, which means that the nonfiler ratio (1 minus the filer ratio) is negative. Because negative values cannot be transformed into logarithms, the log rate equation includes the filer ratio and not the nonfiler ratio. There are several reasons that filer ratios may be greater than 1: addresses on tax returns are not always the county of residence as defined for population estimates; tax filers may report exemptions for children who do not reside with them; and some child exemptions are for children aged 21 or older.

Each single-equation model was estimated for 1993 by averaging 3 years of CPS data (March 1993, 1994, and 1995, covering income years 1992, 1993, and 1994) to form the dependent variable. Each model was also estimated for 1989: for the dependent variable, by averaging 3 years of CPS data (March 1989, 1990, and 1991, covering income years 1988, 1989, and 1990); for the predictor variables, by using appropriate data from IRS and food stamp records for 1989, 1990 population estimates of school-age children, and 1980 census estimates of poor school-age children. The 1989 models were estimated to permit comparisons with 1990 census estimates of poor school-age children in 1989 for evaluation purposes (see Chapter 6). Finally, each single-equation model was also estimated for 1989 by using 1990 census data rather than CPS data to form the dependent variable. The census equation was used to estimate the model error variance of the 1993 and 1989 CPS equations (see Appendix A).

Bivariate Models

The bivariate formulation of the county model for 1993 estimates of poor school-age children involves the joint estimation of two equations: one for 1993, in which the dependent variable is formed by averaging 3 years of CPS data, and one for 1989, in which the dependent variable is formed by using 1990 census data. The bivariate formulation allows for a correlation between the model errors in the two equations (v_{CPSi} and v_{CENi} in equations (2) and (3) below; see also Appendix A). It is through this mechanism that data from the previous census are incorporated in predicting the number of poor school-age children in 1993. Hence, the bivariate models do not include 1990 census estimates of poor school-age children as a predictor variable in the 1993 equation. The bivariate models were not estimated for 1989: a CPS equation for 1989 could have been estimated but a census equation for 1979 could not be estimated because of lack of administrative records data to form predictor variables.

The basic form of the CPS equation in the bivariate formulation is

$$z_{CPSi} = \beta_0 + \beta_1 w_{CPS1i} + \beta_2 w_{CPS2i} \ldots + \beta_4 w_{CPS4i} + v_{CPSi} + a_{CPSi} \ , \qquad (2)$$

where:

z_{CPSi} = the dependent variable in county i (log number or proportion of poor school-age children from 3 years of CPS data),
$w_{CPS1i} \ldots w_{CPS4i}$ = the predictor variables in county i,
v_{CPSi} = model error for county i, and
a_{CPSi} = sampling error of z_{CPSi} for county i.

The basic form of the census equation in the bivariate formulation is

$$z_{CENi} = \beta_0^* + \beta_1^* w_{CEN1i} + \beta_2^* w_{CEN2i} \ldots + \beta_4^* w_{CEN4i} + v_{CENi} + a_{CENi} \quad , \qquad (3)$$

where:

z_{CENi} = the dependent variable in county i (log number or proportion of poor school-age children from the 1990 census),
$w_{CEN1i} \ldots w_{CEN4i}$ = the predictor variables in county i,
v_{CENi} = model error for county i, and
a_{CENi} = sampling error of z_{CENi} for county i.

The formulation with fixed state effects adds an indicator variable for each state, which is 1 for all counties in the state and 0 otherwise.

Six bivariate models were evaluated in the first round (see Table 5-2):

(7) ***Bivariate Log Number Model*** In the CPS equation for this bivariate model, the dependent variable is the same as in model (1), the single-equation log number model (under 21). The predictor variables are the same as in (1), except that the 1990 census estimated number of poor school-age children is dropped from the equation. In the census equation for this bivariate model, the dependent variable is the 1990 census estimate of the log number of poor school-age children in 1989; the predictor variables are the same as in the CPS equation, except that the IRS and food stamp data pertain to 1989 instead of 1993, and the population data are from the 1990 census rather than from the population estimates program. All variables are transformed to logarithms.

(8) ***Bivariate Log Rate Model*** In the CPS equation, the dependent variable is the same as in model (4), the single-equation log rate model (under 21). The predictor variables are the same as in (4), except that the 1990 census estimated poverty rate for school-age children is dropped from the equation. In the 1990 census equation, the dependent variable is the estimated log poverty ratio for school-age children from the census; the predictor variables are the same as in the CPS equation, except that the IRS and food stamp data pertain to 1989 instead of 1993 and the population data are from the 1990 census rather than from the population estimates program. All variables are ratios, transformed to logarithms.

(9) ***Bivariate Rate Model*** The dependent and predictor variables in the CPS and census equations are the same as in (8), but all variables are ratios, untransformed.

(10) ***Bivariate Log Number Model with Fixed State Effects*** The dependent and predictor variables in the CPS and census equations are the same as in (7), with the addition of state indicator variables in each equation. All variables are transformed to logarithms.

(11) ***Bivariate Log Rate Model with Fixed State Effects*** The dependent and predictor variables in the CPS and census equations are the same as in (8), with the addition of state indicator variables in each equation. All variables are ratios, transformed to logarithms.

(12) ***Bivariate Rate Model with Fixed State Effects*** The dependent and predictor variables in the CPS and census equations are the same as in (9), with the addition of state indicator variables in each equation. All variables are ratios, untransformed.

MODELS EXAMINED IN THE SECOND ROUND OF EVALUATIONS

The first round of evaluations included an internal evaluation, in which the regression output for all 12 models was examined to assess the validity of the underlying assumptions (see Appendix B). It also included an external evaluation in which estimates of poor school-age children in 1989 from the six single-equation models were compared with 1990 census estimates (see Appendix C). The results of these evaluations led the Census Bureau and the panel to drop several models from further consideration in the near term.

The untransformed rate model (5) and the hybrid log rate-number model (6) were dropped from consideration because they performed somewhat worse, on balance, than the other models on both the internal and external evaluations. For example, in the comparisons of model estimates of poor school-age children in 1989 with 1990 census estimates, models (5) and (6) exhibited the largest overall absolute differences of their estimates from the census (see Table C-3). Also, the standardized residuals (differences between the model prediction and the reported value for each observation) from the regression equations for models (5) and (6) were not distributed normally.

The bivariate formulation (models 7-12) is promising in that it makes fuller use of the information from the previous census than the single-equation formulation. However, there is less experience with bivariate modeling than with modeling that uses a single equation for the kinds of estimates that are needed. More important, because the IRS and food stamp predictor variables at the county level were not available for 1979, it is not possible to evaluate bivariate models by comparison with estimates from the 1990 census. (Such a model would require joint estimation of a 1989 equation in which CPS data form the dependent variable and a 1979 equation in which 1980 census data form the dependent variable.) Hence, the bivariate formulation was not pursued for use in the short run. However, further development of bivariate and multivariate models, which might include CPS equations for more than 1 year, as well as a census equation, is worth pursuing for the longer run (see Chapter 9).

Evaluation results indicated that the county model would likely benefit from taking account of state effects in some way. The addition of state indicator

TABLE 5-2 Bivariate County Models: Dependent Variable, Predictor Variables, and Form of the Predictor Variables for the CPS Equation for 1993

Model	Dependent Variable, z_{CPSi}	Predictor Variables, $w_{CPS1i} \ldots w_{CPS4i}$	Form of the Predictor Variables
(7) Log Number (Under 21)	Log 3-year weighted average number of poor school-age children (same as single-equation Log Number Under 21 (1))	(1) Number of child exemptions reported by families in poverty on tax returns (2) Number of people receiving food stamps (3) Population under 21 (4) Number of child exemptions on tax returns (same as single-equation Log Number Under 21, except there is no previous census variable)	Transformed to Logarithms
(8) Log Rate	Log poverty ratio for school-age children (3-year sum of poor related children aged 5-17 divided by 3-year sum of total CPS children aged 5-17) (same as single-equation Log Rate (4))	(1) Ratio of number of child exemptions reported by families in poverty on tax returns to total number of child exemptions on tax returns (2) Ratio of number of people receiving food stamps to total population (3) Ratio of total number of child exemptions on tax returns to total population under 21 (same as single-equation Log Rate, except there is no previous census variable)	Transformed to Logarithms

(9) Rate	Poverty ratio for school-age children (same as Bivariate Log Rate (8), except untransformed)	Same as Bivariate Log Rate, except untransformed	Untransformed
(10) Log Number with Fixed State Effects	Same as Bivariate Log Number Under 21 (7)	Same as Bivariate Log Number Under 21 with the addition of state indicator variables	Transformed to Logarithms
(11) Log Rate with Fixed State Effects	Same as Bivariate Log Rate (8)	Same as Bivariate Log Rate with the addition of state indicator variables	Transformed to Logarithms
(12) Rate with Fixed State Effects	Poverty ratio for school-age children (same as Bivariate Log Rate (8), except untransformed)	Same as Bivariate Log Rate with the addition of state indicator variables, except untransformed	Untransformed

NOTES: The models are estimated for 1993 from a CPS equation for 1993 and a 1990 census equation for 1989. The census equation for 1989 for each bivariate model is of the same form as the corresponding CPS equation for 1993. The 1989 equations use the number of poor school-age children or the poverty ratio for school-age children from the 1990 census as the dependent variable; the predictor variables are from IRS and food stamp records for 1989 and population estimates from the 1990 census.

variables to either a single-equation or bivariate model (3, 10-12) was promising in some respects, but a fixed state effects approach did not seem clearly superior to other models that were examined. There was not time to investigate other approaches to account for state effects, although the panel believes that the county model might be improved in this regard with more research (see Chapter 9).

At the conclusion of the first round of evaluations, the Census Bureau and the panel focused on four models that were considered serious candidates to produce revised 1993 county estimates of poor school-age children. These four candidate models were then evaluated on several criteria. All four models are of the single-equation form with variables transformed to logarithms and without fixed state effects:

(a) Log number model (under 21), model (1) above, used by the Census Bureau to produce the original 1993 county estimates of poor school-age children.

(b) Log number model (under 18), model (2) above. This model is the same as model (a) except that the population under age 18 replaces the population under age 21 as a predictor variable.

(c) Log rate model (under 21), model (4) above. The rate formulation is used in the Census Bureau's state model, and the panel believed that, in log form, it might improve the county model.

(d) Log rate model (under 18). This model is the same as model (c) except that the ratio of total child exemptions on tax returns to the total population under 18 replaces the ratio of total child exemptions on tax returns to the total population under age 21 as a predictor variable. The panel wanted to determine if this modification would improve the log rate model, since a similar modification had been found to improve the log number model. However, for reasons that are not clear, this modification to the log rate model worsened rather than improved its performance in several respects (see Chapter 6).

The model that the Census Bureau used to prepare the revised 1993 county estimates of poor school-age children is (b)–log number model (under 18), estimated with maximum likelihood. This model was also used to prepare the 1995 estimates. Chapter 6 describes the evaluations that were conducted of the four candidate models (a-d) and highlights key results. Appendix B analyzes the regression output for the 12 models that were included in the first round of evaluations and model (d). Appendix C provides 1990 census evaluation results for the six single-equation models that were included in the first round of evaluations and the four candidate models that were evaluated in the second round. Appendix C also compares the four candidate models with four other procedures that rely more heavily on census data. These procedures are described and discussed in Chapter 6.

6

Evaluations of County Estimates

The development of model-based estimates for small areas is a major, continuing research and development effort for which extensive evaluation is required. For updated estimates of poor school-age children for counties, a thorough assessment of all aspects of the estimation procedure is necessary to have confidence in the estimates–whether the estimates are used by the Department of Education to allocate Title I funds to counties (as was the practice before the 1999-2000 school year) or whether they are used to develop estimates for school districts.

The Census Bureau's county estimates of poor school-age children are produced by using a county regression model and a state regression model (see Chapter 4).[1] A comprehensive evaluation of these two components of the estimation procedure should include both "internal" and "external" evaluations.

The first test of a regression model is that it perform well when evaluated internally, that is, for the set of observations for which it is estimated. Such an internal evaluation is primarily an investigation of the validity of the model's underlying assumptions and features, which for a regression model is typically based on an examination of the residuals from the regression–the differences between the predicted and reported values of the dependent variable for each observation.

[1]Population estimates of school-age children are provided to accompany the estimates of poor school-age children to permit calculating poverty rates–see Chapter 8 for a description of the methods used for postcensal population estimates and for evaluation results.

In an external evaluation, the estimates from a model are compared with target or "true" values that were not used to develop the model. Ideally, an internal evaluation of regression model output should precede external evaluation. Changes made to the model to address concerns raised by the internal evaluation would likely improve its performance in the external evaluation.

Since there are no absolute criteria for what are acceptable evaluation results, one method for determining if the performance of a model can be improved is to examine alternative models. Such comparisons may indicate changes that would be helpful for a model; they may also suggest that an alternative model is preferable. Both internal and external evaluations should be carried out for alternative models.

OVERVIEW OF EVALUATIONS

1993 Estimates

When the original 1993 county estimates of poor school-age children were provided to the panel, the Census Bureau had not had time to complete a full evaluation of them. Subsequently, the panel developed a set of evaluation criteria, and the panel and the Census Bureau conducted a series of internal and external evaluations. The focus of the evaluation effort was on alternative county models, particularly the assumptions underlying the regression equations and how the estimates of poor school-age children in 1989 from each model compared with 1990 census estimates. The state model was examined as well, both directly and as it contributed to the county estimates of poor school-age children. The evaluations included:

(1) internal evaluation of the regression output for alternative county models estimated for 1993 and 1989;
(2) comparison of estimates of poor school-age children for 1989 from alternative county models with 1990 census estimates, a form of external evaluation;
(3) examination of the original 1993 county estimates to identify possibly anomalous estimates that were then reviewed with knowledgeable local people, another form of external evaluation; and
(4) evaluation of the state model, including examination of regression output and external evaluation in comparison with 1990 census estimates.

The internal evaluation of regression output and the comparison of model-based estimates of poor school-age children for 1989 with 1990 census estimates–evaluations (1) and (2) above–were carried out for the four single-equation county models that were considered serious candidates to produce revised 1993 county estimates of poor school-age children (see Chapter 5 and Appendices B and C):

(a) log number model (under 21), the original model that the Census Bureau used to produce the original 1993 county estimates of poor school-age children;
(b) log number model (under 18), the revised model that the Census Bureau used to produce the revised 1993 county estimates of poor school-age children;
(c) log rate model (under 21); and
(d) log rate model (under 18).

In addition, the 1990 census comparisons (2) were performed for some other estimation procedures that relied much more heavily than did the four candidate models on estimates from the 1980 census (see below, "Comparisons with 1990 Census Estimates"). Since the Department of Education used estimates of poor school-age children from the previous census for allocations of Title I funds prior to the 1997-1998 school year, these estimation procedures were included in the evaluation in order to see how well the regression models compared with some simple procedures for updating the census estimates.

The internal evaluation of regression output (1) and the comparison of estimates of poor school-age children for 1989 with 1990 census estimates (2) examined residuals and model differences from the census, respectively, for categories of counties. The following characteristics were used for categorizing counties: census geographic division; metropolitan status of county; population size in 1990; population growth from 1980 to 1990; percentage of poor school-age children in 1980; percentage of Hispanic population in 1990; percentage of black population in 1990; persistent poverty from 1960 to 1990 for rural counties; economic type for rural counties; percentage of group quarters residents in 1990; number of households in the CPS sample in 1988-1991 (or whether the county had sampled households); and (for 1990 census comparisons only) percentage change in the poverty rate for poor school-age children from 1980 to 1990 (see details in Table 6-4, below).

1995 Estimates

Because the 1995 county estimates were developed by using a procedure similar to that used to develop the revised 1993 county estimates, the focus of the evaluation effort for the 1995 estimates shifted to how the state and county models behaved over several time periods, and specifically, to determining whether there were persistent biases or other problems. The evaluations of the 1995 county estimates included:

(1) internal evaluation of the regression output for the 1995 county model estimated for 1995, 1993, and 1989 (using uncorrected and corrected tax return data);

(2) comparison of estimates of poor school-age children that were developed from the 1995 form of the county model for 1995, 1993, and 1989 with CPS estimates for groups of counties, a form of external evaluation; and

(3) evaluation of the state model, including examination of regression output for 1996, 1995, 1993, 1992, 1991, 1990, and 1989 and consideration of the state raking factors by which county model estimates are adjusted to make them consistent with the state model estimates.

COUNTY MODEL INTERNAL EVALUATION

1993 Evaluations

The panel and the Census Bureau examined the underlying assumptions and other features of the four models, (a)-(d), that were considered candidates for producing revised 1993 county estimates of poor school-age children, through evaluation of the regression model output for 1989 and 1993.[2] Although such an evaluation is not likely to provide conclusive evidence with which to rank the performance of alternative models, particularly when they use different transformations of the dependent variable, examination of the regression output is helpful to determine which models perform reasonably well.

The assumptions and features investigated for the four models fall into two groups: those concerning the functional form of the regression model and those concerning the error distribution. Because properties of the error distribution affect the ability to fit a model, studies of these two types of assumptions are not entirely separable.[3]

The assumptions and features examined in the first group are linearity of the relationship between the dependent variable and the predictor variables; constancy of the assumed linear relationship over different time periods; and whether

[2]The evaluation of the county regression output pertains to the regression models themselves, that is, before the predictions are combined with the direct CPS estimates in a "shrinkage" procedure or raked to the estimates from the state model (see Chapter 4). For these models, the regression output comprises the model predictions for counties with at least one household with poor school-age children in the CPS sample. For the two log number models, the predictions are the log number of poor school-age children; for the two log rate models, the predictions are the log proportion of poor school-age children.

[3]These assumptions were also examined for the analogous 1990 census regressions. However, since the census equations only affected the weights for the weighted least squares regression and the extent of "shrinkage" in combining model estimates and direct estimates for counties with households in the CPS sample, analyses of the 1990 census regressions are not discussed here.

any of the included predictor variables are *not* needed in the model and, conversely, whether other potential predictor variables *are* needed in the model. The assumptions examined in the second group are normality (primarily symmetry and moderate tail length) of the distribution of the standardized residuals;[4] whether the standardized residuals have homogeneous variances, that is, whether the variability of the standardized residuals is constant across counties and does not depend on the values of the predictor variables; and absence of outliers. Each assumption is discussed in terms of the methods used for evaluation and the results of the evaluation for the four candidate models.

Linearity of the relationships between the dependent variable and the predictor variables was assessed graphically, by observing whether there was evidence of curvature in the plots of standardized residuals against the predictor variables in the model. In addition, plots of standardized residuals against CPS sample size and against the predicted values from the regression model were also examined for curvature.

The only evidence of nonlinearity is for the log number (under 21) model (a) for 1989. For that year, the standardized residuals appear to have a very modest curvature when plotted against the predicted values.

Constancy over Time of the assumed linear relationship of the dependent and predictor variables was assessed through comparison of the regression coefficients on the predictor variables for 1989 and 1993. While major changes in economic conditions are expected to cause some changes in the coefficients, a relatively stable regression equation would be desirable.

Table 6-1 shows the regression coefficients for the predictor variables for the four candidate models for 1989 and 1993. In the log number models (a, b) for 1989 and 1993, the coefficients for the three "poverty level" predictor variables—child exemptions reported by families in poverty on tax returns (column 1), food stamp recipients (column 2), and poor school-age children from the previous census (column 5)—are similar. There are substantial differences across the two time periods in the estimated coefficients for the other two variables—population (under age 21 or under age 18, column 3) and total number of child exemptions on tax returns (column 4). However, the sum of these two coefficients is generally close to 0 in each model in each year. Because these two variables are highly positively correlated, the predictions from equations with a similar sum for the two coefficients will be similar.

[4]The standardization of the residuals involved estimating the predicted standard errors of the residuals, given the predictor variables, and dividing the observed residuals by the predicted standard errors. The predicted standard error of the residual for a county is a function of the estimated model error variance and the estimated sampling error variance (see Belsley, Kuh, and Welsch, 1980).

TABLE 6-1 Estimates of Regression Coefficients for Four Candidate County Models for 1989 and 1993

Model	Counties (Number)	Predictor Variables[a] 1	2	3	4	5
(a) Log Number (under 21)						
1989	1,028	0.52	0.30	0.76	–0.81	0.27
		(.07)	(.05)	(.22)	(.22)	(.07)
1993	1,184	0.31	0.30	0.03	0.03	0.40
		(.08)	(.07)	(.21)	(.21)	(.09)
(b) Log Number (under 18)						
1989	1,028	0.50	0.23	1.79	–1.80	0.32
		(.06)	(.05)	(.27)	(.27)	(.07)
1993	1,184	0.38	0.27	0.65	–0.59	0.34
		(.08)	(.07)	(.24)	(.24)	(.09)
		Predictor Variables[b]				
(c) Log Rate (under 21)						
1989	1,028	0.32	0.29	–0.73	0.40	
		(.07)	(.04)	(.19)	(.07)	
1993	1,184	0.23	0.31	–0.07	0.41	
		(.08)	(.06)	(.18)	(.09)	
(d) Log Rate (under 18)						
1989	1,028	0.29	0.26	–1.13	0.43	
		(.07)	(.04)	(.24)	(.07)	
1993	1,184	0.26	0.30	–0.42	0.38	
		(.08)	(.06)	(.20)	(.09)	

NOTES: All predictor variables are on the logarithmic scale for numbers and rates. Standard errors of the estimated regression coefficients are in parentheses. The four models were estimated for each year with maximum likelihood. The original 1994 population estimates were used for the 1993 models; 1990 census population estimates were used for the 1989 models.

[a]Predictor variables: (1) number of child exemptions reported by families in poverty on tax returns; (2) number of people receiving food stamps; (3) population (under age 21 or under age 18); (4) total number of child exemptions on tax returns; (5) number of poor school-age children from previous (1980 or 1990) census.

[b]Predictor variables: (1) ratio of child exemptions reported by families in poverty on tax returns to total child exemptions; (2) ratio of people receiving food stamps to total population; (3) ratio of total child exemptions on tax returns to population (under age 21 or under age 18); (4) ratio of poor school-age children to total school-age children from previous (1980 or 1990) census.

The sum of all coefficients in each equation for models (a) and (b) ranges from 1.04 to 1.07 and is significantly greater than 1. A sum equal to 1 would mean that county population size itself has no effect on the estimated number of poor school-age children and that the model is expressible as a model with the poverty rate as the dependent variable and rates as predictor variables. Because the sum is greater than 1, the estimated number of poor school-age children is a larger percentage of the population in the larger counties. While this result is difficult to explain as a function of county size, it may be that size reflects the effects of variables not included in the models.

In the log rate models (c, d), the coefficients for the three "poverty rate" predictor variables—ratio of child exemptions reported by families in poverty on tax returns to total child exemptions (column 1), ratio of food stamp recipients to the total population (column 2), and ratio of poor school-age children to total school-age children from the previous census (column 4)—are all positive and about the same size.[5] The coefficients for the ratio of total child tax exemptions to the population (under age 21 or under age 18, column 3) are negative, as is also generally the case for the coefficients of the related variable (total number of child tax exemptions) in the log number equations. There are substantial differences in the estimated coefficients for the ratio of total child tax exemptions to the population in the log rate models across time periods and some differences between the coefficients in the two models.

Inclusion or Exclusion of Predictor Variables The possibility that one or more predictor variables should be excluded from a model was assessed by looking for insignificant *t*-statistics for the estimated values of individual regression coefficients.[6] The need to include a predictor variable, or possibly to model some categories of counties separately, was assessed by looking for nonrandom patterns, indicative of possible model bias, in the distributions of standardized residuals displayed for the various categories of counties.[7]

The only predictor variables with nonsignificant *t*-statistics are the population under age 21 (column 3 in Table 6-1) and total child exemptions on IRS income tax returns (column 4) for the log number (under 21) model (a) in 1993, and the ratio of child tax exemptions to the population under age 21 (column 3) for the log rate (under 21) model (c) in 1993. All other regression coefficients are

[5]The coefficients are also similar to the coefficients for the corresponding variables—number of child exemptions reported by families in poverty on tax returns, number of food stamp recipients, and number of poor school-age children from the previous census—in the log number equations.

[6]Although the performance of a predictive regression model is best assessed in terms of the joint impact of the predictor variables, examining the individual predictor variables can suggest ways in which a model might be improved.

[7]The distributional displays examined for this and other model assumptions were box plots.

significantly different from 0 at the 5 percent level. Application of Akaike's information criterion (AIC) confirmed the superiority of using the population under age 18 as a predictor variable in preference to the population under age 21 in the log number model. (The test was not performed for the log rate model.)

For most ways of categorizing counties, the standardized residuals do not exhibit systematic patterns. The exceptions are that all four models in 1989 tend to overpredict poor school-age children in counties with a high percentage of Hispanic residents (i.e., the standardized residuals tend to be negative for these counties) and that the log number (under 21 and under 18) models (a, b) in 1993 and 1989 tend to overpredict poor school-age children in counties that are in metropolitan areas but are not the central county in the area.

Normality of the standardized residuals was evaluated through use of Q-Q plots, which match the observed distribution of the residuals with the theoretical distribution, and other displays of the distribution. All four models exhibit some skewness in their standardized residuals, with the log rate models (c, d) showing somewhat more skewness than the log number models (a, b). For none of the models does the skewness appear sufficiently marked to be a problem.

Homogeneous Variances The homogeneity of the variance of the standardized residuals was assessed using a variety of statistics and graphical displays (see Appendix B). Examination of them clearly demonstrates some variability in the size of the absolute standardized residuals as a function of the predicted value (number or proportion of poor school-age children) and the CPS sample size for all four models. With regard to CPS sample size, one would expect the standardized residual variance to remain constant over the distribution of CPS sample size; however, it increases with increasing CPS sample size.

The heterogeneity of the variance of the residuals suggests that there may be a problem with the model specification or in the assumptions that were used to calculate the standardized residuals. However, adjusting a model to remove this type of heterogeneity is likely to have only a small effect on the estimated regression coefficients or the model estimates. The effect on estimates of poor school-age children would stem from two factors: a shift in the weights assigned to each county in fitting the regression model, which would very likely result in only a modest change in the estimated regression coefficients; and a change in the weight given to the direct estimates, which could have an appreciable effect on the estimates only for the few counties with large CPS sample sizes.

Outliers The existence of outliers was evaluated through examination of plots of the distributions of the standardized residuals and plots of standardized residuals against the predictor variables and through analysis of patterns in the distribution of the 30 largest absolute standardized residuals for the various categories of counties. However, it is difficult to evaluate the evidence for outliers that results from a least squares model fit, which has the property that it may miss influential outliers. In addition, since the four models are so similar and make use of the identical data, it is unlikely that an observation that was a marked outlier for one model would not also be a marked outlier for the other models.

An examination of the distributions of the standardized residuals suggests that none of the four models is especially affected by outliers, although the 1993 models have more outliers than the 1989 models, and nonrural counties and metropolitan counties that are not central counties have somewhat more outliers than other categories of counties. This analysis is only a start. It would be useful to extend this analysis, using other statistics and various graphical techniques, to identify the counties that are not well fit by robustly estimated versions of these models in order to determine any characteristics that outlier counties have in common.

Summary The panel concluded that the analysis of the regression output for the four candidate county models for 1989 and 1993 largely supports the assumptions of the models: there is little evidence of important problems with the assumptions. The analysis does not strongly support one model over another, although it does support use of the population under age 18 instead of the population under age 21 as a predictor variable in the log number model.

All of the models exhibit a few common problems. First, they all behave somewhat differently for larger urban counties and counties with large percentages of Hispanic residents than for other counties. Second, all models show evidence of some variance heterogeneity with respect to both CPS sample size and the number or proportion of poor school-age children.

1995 Evaluations

The internal evaluation for the 1995 county model, which is essentially the log number (under 18) model (b) evaluated above, focused on comparisons of the properties of the model when estimated for different time periods. The analysis looked in particular at three characteristics: the constancy of the regression coefficients for the predictor variables over time; distributions (box plots) of the standardized residuals for categories of counties to determine if there were any nonrandom patterns that persisted over time; and the phenomenon observed in the 1993 evaluations by which the variance of the standardized residuals was related to CPS sample size and the predicted value of the dependent variable (variance heterogeneity).

Constancy of the Regression Coefficients Because the county model is refitted for each prediction year, constancy of the regression coefficients for the predictor variables over time is not as important as it would be if the estimated regression coefficients from the model were used for predictions for subsequent

TABLE 6-2 Estimates of Regression Coefficients for Census Bureau 1995 County Model, Estimated for 1989, 1993, and 1995

		Predictor Variables[a]				
Year	No. of Counties	(1)	(2)	(3)	(4)	(5)
1989 (revised IRS data)	1,028	0.52 (.06)	0.29 (.06)	1.55 (.31)	–1.56 (.30)	0.26 (.06)
1989 (original IRS data)	1,028	0.50 (.06)	0.23 (.05)	1.79 (.27)	–1.80 (.27)	0.32 (.07)
1993	1,184	0.38 (.08)	0.27 (.07)	0.65 (.24)	–0.59 (.24)	0.34 (.09)
1995	985	0.31 (.10)	0.29 (.08)	0.88 (.25)	–0.80 (.25)	0.33 (.09)

NOTE: All predictor variables are on the logarithmic scale for numbers. Standard errors of the estimated regression coefficients are in parentheses.

[a]Predictor variables: (1) number of child exemptions reported by families in poverty on tax returns; (2) number of people receiving food stamps; (3) population under age 18; (4) total number of child exemptions on tax returns; (5) number of poor school-age children from previous (1980 or 1990) census.

years. Also, major changes in economic conditions would be expected to cause some changes in the coefficients. Nonetheless, it is desirable for the coefficients to be in the same direction and not fluctuate wildly in size over time.

Table 6-2 shows the regression coefficients for the predictor variables for the 1995 county model estimated for 1995 and 1993 and for 1989 with both the original and revised IRS data (see Chapter 4).[8] The coefficients for the three "poverty level" predictor variables—child exemptions reported by families in poverty on tax returns (column 1), food stamp recipients (column 2), and poor school-age children from the previous census (column 5)—are fairly similar in the equations for all three time periods. There are more substantial differences across the three time periods in the size of the estimated coefficients for the other two variables—population under age 18 (column 3) and total number of child exemptions on tax returns (column 4). However, the sum of these two coefficients is close to zero in each year. Because the two variables are highly posi-

[8]The regressions for 1995 and for 1989 with corrected IRS data also used modified food stamp data (i.e., the county food stamp data were raked to the adjusted state food stamp data, as described in Chapter 4).

tively correlated and close in magnitude, the predictions from equations with a similar sum for the two coefficients will be similar.

Finally, the sum of all the coefficients is close to 1 for all 3 estimation years: 1.01 for 1995, 1.05 for 1993, and 1.06 for 1989 with the revised IRS data. It is desirable for the coefficients in a model of this form to sum to 1, which indicates that the model predictions do not vary by the scale of the predictor variables. If the sum of the coefficients is much greater than or less than 1, the model should be examined to determine if additional predictor variables or other changes in the model may be needed.

Patterns of Residuals Given typical random variation, it is likely that the distributions of standardized residuals will display apparently nonrandom patterns for some categories of counties in a particular year. However, if the distributions display the same patterns across years, it is evidence of model bias. The persistence of the same patterns should be investigated to determine ways to eliminate or reduce the bias, for example, by adding a variable to the equation. (There are ample degrees of freedom in the county model to permit the inclusion of additional predictor variables.)

Investigation of the standardized residuals for categories of counties for the county model estimated for 1995, 1993, and 1989 reveals little evidence of persistent bias. However, there is some suggestion that the model tends to consistently overpredict the number of poor school-age children in smaller size counties (i.e., the model estimates are somewhat higher than the CPS direct estimates for smaller counties). It also tends to overpredict the number of poor school-age children in counties that are in metropolitan areas but are not the central county in the area. These patterns, while not strong, are evident in the regression output for all 3 years. The tendency for the model to overpredict the number of poor school-age children in counties with a high percentage of Hispanics that was evident for 1989 in the 1993 model evaluations did not persist over time.

Variance Heterogeneity The regression output for the 1995 county model clearly demonstrates variability in the size of the absolute standardized residuals as a function of the predicted value (log number of poor school-age children) and the CPS sample size. If the variance estimates for the model are correct, then the standardized residual variance should remain constant over the distribution of CPS sample size. However, it increases with increasing CPS sample size. This phenomenon was evident in the evaluations conducted for the 1993 county model, and it is evident in all 3 years for which the 1995 county model was estimated.

As noted for the 1993 evaluations above, adjusting a model to remove this type of heterogeneity is likely to have only a small effect on the estimated regression coefficients or the model estimates (although it will affect the estimated confidence intervals around the model estimates). Nonetheless, it is clear that the current method for estimating the variance of the sampling errors—a_i in equation

(1) in Chapter 4—in the county model is incorrect. The current approach estimates the model error variance from a 1989 equation in which 1990 census data form the dependent variable, and then uses the estimate for the model error variance in the CPS-based county equation (see Chapter 4). Taking this estimated model error variance as fixed, the total sampling error variance is obtained together with estimated regression coefficients using a maximum likelihood procedure. Finally, the total sampling error variance is distributed to counties by assuming that the sampling error variance in a county is inversely proportional to the county's CPS sample size. An alternative approach for estimating the sampling error variance that might remove the variance heterogeneity in the regression residuals is discussed in Chapter 9 (see also National Research Council, 2000:Ch.3).

Summary The panel concluded that the analysis of the regression output for the 1995 county model estimated for 1989, 1993, and 1995 largely supports the assumptions of the model: there is little evidence of important problems with the assumptions. However, the model does exhibit a few minor problems that appear to persist over time. First, it tends to overpredict the number of poor school-age children in smaller counties and metropolitan counties that are not the central county. The differences are not marked, but research should be conducted to determine possible ways to modify the model to eliminate or reduce this problem. Second, the model shows evidence of variance heterogeneity with respect to both CPS sample size and predicted number of poor school-age children. Improvements in estimating the model error and sampling error variances should be sought to reduce or eliminate this problem.

COUNTY MODEL EXTERNAL EVALUATION

Comparisons with 1990 Census Estimates

For external evaluation of alternative models that were considered for 1993 estimates, the panel and the Census Bureau compared the estimated number and proportion of poor school-age children for 1989 for the four candidate models with 1990 census estimates.[9] The evaluation examined the overall difference

[9]The county estimates reflect the effects of the state model and the county population estimates as well as the county regression model, but the differences in model performance vis-à-vis the census in the evaluation are due to the particular form of the county model.

The models for which the 1990 census comparisons were performed were estimated with the method of moments. Maximum likelihood was used to estimate the log number (under 18) model (b) for the revised 1993 county estimates and the 1995 county estimates of poor school-age children. The differences in the estimates from the two techniques are small.

between the estimates from a model and the census and the differences for groups of counties categorized by various characteristics.

Evaluation by comparison with the 1990 census is not ideal because the census estimates are not true values. They are affected by sampling variability and population undercount; also, the census measurement of poverty differs from the CPS measurement in ways that are not fully understood (see Chapter 3). In addition, there is only one census-based validation opportunity: because of the lack of IRS and Food Stamp Program data for counties for 1979, it is not possible to evaluate model-based estimates by comparison to the 1980 census. Reliance on a single validation using the 1990 census is a problem because a model may perform better or worse in any one validation than it would on average over multiple validations. For this reason, if it were possible to compare model estimates with census or other estimates for 1993 instead of 1989, the results might turn out differently. Nonetheless, in the absence of other means of external validation, the panel and the Census Bureau relied heavily on the 1990 census comparisons to understand the performance of alternative models.

Evaluation by comparison with the 1990 census is intended to assess the accuracy of model estimates for the prediction year (i.e., 1989). The evaluation does not address the issue that model-based estimates for a given year are used for Title I allocations about 3 years later.

The 1990 census estimates that are used in the comparisons are ratio adjusted by a constant factor to make the census national estimate of poor school-age children equal the 1989 CPS national estimate. This adjustment removes the difference of about 6 percent between the CPS and census estimates of total poor school-age children for 1989. Consequently, the differences between a model and the 1990 census in estimating poor school-age children for groups of counties can be interpreted as differences in shares. This feature is useful because the Title I allocation formula distributes funding as shares (percentages) of a fixed total dollar amount.

In addition to the four candidate models, the 1990 census comparisons were performed for four estimation procedures that rely much more heavily on 1980 census estimates. Given the substantial changes in the number and proportion of poor school-age children between the 1980 and 1990 censuses (see Chapter 3), one would expect these procedures to perform less well than the candidate models in predicting poverty for school-age children in 1989.[10] In a period of less pronounced change, one or more of them might perform relatively well. The census comparisons were done for the following procedures:

[10]Although the interval was only 4 years instead of 10, substantial changes in the number and proportion of poor school-age children also occurred between 1989 and 1993, and such changes continued to be observed through 1999 (see Chapter 3).

(i) Stable shares procedure, in which the county estimates of poor school-age children for 1989 are the 1980 census estimates for 1979 after ratio adjustment to make the 1980 census national estimate equal the CPS national estimate for 1989. This simple procedure assumes no change over the decade in each county's share of the total number of poor school-age children nationwide: this is the same assumption that underlies previous practice for Title I allocations, in which estimates from the decennial census were used in the formulas each year until the results from the next census became available.[11]

(ii) Stable shares within state procedure, in which the county estimates of poor school-age children for 1989 are the 1980 census estimates for 1979 after raking the estimates for the counties in each state to the estimates from the Census Bureau's state model for 1989. (The national raking employed in the state model also adjusts the total to equal the CPS national estimate for 1989.) This procedure assumes no change over the decade in each county's share of the total number of poor school-age children in its state.

(iii) Stable rates within state procedure (with conversion), in which the county estimates of poor school-age children for 1989 are developed by converting 1980 census estimates of the proportions of poor school-age children for 1979 to estimated numbers by use of 1990 county population estimates of total school-age children 5-17 and then raking the estimated numbers to the Census Bureau's state model estimates for 1989.

(iv) Averaging procedure, in which the county estimates of poor school-age children for 1989 are developed from an average of estimates from the 1980 census and the log number (under 21) model (a) for 1989.[12]

The rest of this section first discusses overall absolute differences from the 1990 census estimates for the four candidate models and the four procedures that rely more heavily on the 1980 census. It then discusses differences for categories of counties for the four candidate models and two of the procedures: the stable shares procedure and the averaging procedure. Differences for categories of

[11]However, the estimates from the 1990 census that were previously used for Title I allocations were not adjusted to the current CPS national estimate of poor school-age children, which could affect the allocations for some counties. For example, some counties might meet the threshold test for a concentration grant if the census estimates were adjusted to the current CPS national estimate but not if the estimates were unadjusted.

[12]More precisely, the estimates are developed by averaging the proportions of poor school-age children from the 1980 census and the log number (under 21) model (a) for 1989, converting the estimates to numbers by the use of 1990 county population estimates of total school-age children, and making an overall ratio adjustment to the CPS national estimate for 1989.

This procedure is analogous to the panel's recommendation for averaging 1990 census and 1993 model-based estimates for use in Title I allocations for the 1997-1998 school year. However, the panel's recommendation did not include raking the average estimates to the CPS national estimate of poor school-age children in 1993 (see National Research Council, 1997:38).

counties for the other two procedures, which are intermediate in their reliance on 1980 census estimates, are provided in Appendix C.

Absolute Differences Between Model and Census County Estimates

Table 6-3 presents measures of the overall absolute difference between the model-based county estimates and the 1990 census county estimates of poor school-age children in 1989 for the four candidate models and the four procedures that rely more heavily on the 1980 census. If the 1990 census estimates are reasonably accurate, a good model will produce estimates that differ little from the census estimates, and the absolute differences will be less than for other reasonable models. Also, a good model will perform significantly better than a simple procedure that relies heavily on the previous census.

Column 1 of Table 6-3 is the average absolute difference for county estimates of the number of poor school-age children in 1989, measured as the sum for all counties of the absolute difference (ignoring the direction of the difference) between the model estimate and the 1990 census estimate for each county, divided by the total number of counties. Column 2 of Table 6-3 is the average proportional absolute difference for county estimates of the number of poor school-age children, measured as the sum for all counties of the absolute difference between the model estimate and the 1990 census estimate as a *proportion* of the census estimate for each county, divided by the total number of counties and expressed as a percentage. Column 3 is the average proportional absolute difference for county estimates of the proportion of poor school-age children. Column 3 is of interest because the proportion of poor school-age children is used as an eligibility threshold for Title I grants.

The measure in column 1 assesses the difference between a model and the 1990 census in terms of numbers of poor children; the measures in columns 2 and 3 assess the difference in terms of percentage errors for counties. To illustrate the difference between absolute and proportional absolute differences, consider two counties, one with an estimated 10,000 poor school-age children from the census and an estimated 9,600 poor school-age children from the model and the other with an estimated 1,000 poor school-age children from the census and an estimated 1,400 poor school-age children from the model. The absolute difference in the number of poor school-age children is the same for both counties (400), but the proportional absolute difference is only 4 percent for the first county and 40 percent for the second.

From a national perspective, it can be argued that absolute differences are more important for effective Title I allocations because Title I funds are primarily distributed in proportion to the number of children in a county; therefore, the amount of funds that are misallocated depends primarily on the number of children rather than the percentages by county. For example, an error of 5 percent in the number of school-age children in poverty in a large county could correspond

TABLE 6-3 Comparison of Model Estimates and Other Procedures with 1990 Census County Estimates of the Number and Proportion of Poor Related Children Aged 5-17 in 1989

	Average Absolute Difference	Average Proportional Absolute Difference, in Percent	
Model	1 Number of Poor Children Aged 5-17[a]	2 Number of Poor Children Aged 5-17[b]	3 Proportion of Poor Children Aged 5-17[c]
Candidate Models			
(a) Log number (under 21)	272	15.4	16.4
(b) Log number (under 18)	268	16.4	17.7
(c) Log rate (under 21)	275	17.5	17.1
(d) Log rate (under 18)	283	18.8	18.6
Procedures that Rely More Heavily on the 1980 Census			
(i) Stable shares	570	30.1	N.A.
(ii) Stable shares within state	380	27.1	N.A.
(iii) Stable rates within state, with conversion	381	26.2	N.A.
(iv) Average of 1980 census and 1989 log number (under 21) model (a)	286	19.0	N.A.

NOTES: The census estimates are controlled to the CPS national estimate for 1989. See text for definitions of models and measures; N.A.: not available.

[a]The formula where there are n counties (i), is $\Sigma(|Y_{\text{model } i} - Y_{\text{census } i}|) / n$.
[b]The formula is $\Sigma\ [(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}] / n$.
[c]The formula is $\Sigma\ [(|P_{\text{model } i} - P_{\text{census } i}|) / P_{\text{census } i}] / n$.

SOURCE: Data from U.S. Census Bureau.

to tens of thousands of children and have more impact on the allocation of funds than errors of 5 percent in several smaller counties. However, from the county perspective, proportional errors are also important. Ideally, a model will perform well on both types of measures.

The panel drew several conclusions from Table 6-3:

- The performance of the four candidate models is similar, which is not surprising, given that they are variations of the same basic formulation. Thus, the range of the average absolute difference in the estimated number of poor school-age children (column 1) is from 268 children (model b) to 283 children (model d). The average county had about 2,500 poor school-age children for 1989, so that the average absolute difference ranges from 10.7 to 11.3 percent. The range of the average proportional absolute difference in the estimated number of poor school-age children (column 2) is somewhat larger, from 15.4 percent (model a) to 18.8 percent (model d).
- The log number models (a, b) have somewhat lower average absolute differences for estimates of numbers of poor school-age children than do the log rate models (c, d). This is expected because the estimates from the log rate models must be converted to numbers by use of population estimates of total school-age children, which themselves contain error (see Chapter 8). It was expected for the same reason that the log number models would have higher average absolute differences for estimates of proportions of poor school-age children than would the log rate models because population estimates must be used to convert the estimated numbers from the log number models to estimated proportions. However, model (a) shows lower and model (b) shows not appreciably higher average proportional absolute differences for estimates of poverty rates compared with the better log rate model (c)—see column 3 of Table 6-3.
- The four candidate models substantially outperform the three procedures (i-iii) that rely solely or largely on 1980 census data. For example, the *largest* average absolute difference for the four candidate models is 283 poor school-age children (11% of the average number) for the log rate (under 18) model (d), while the *smallest* average absolute difference for procedures (i-iii) is 380 poor school-age children (15% of the average number) for the procedure that assumes stable poverty shares within state (ii). The differences are even somewhat larger for the average proportional absolute difference for estimates of the number of poor school-age children: 18.8 percent for the worst candidate model, model (d), compared with 26.2 percent for the best procedure of these three, the procedure that assumes stable poverty rates within state with conversion (iii).
- The four candidate models also perform better than the procedure (iv) that averages 1980 census estimates with estimates from the log number (under 21) model (a) for 1989, although the differences are not large.

Category Differences in Numbers of Poor School-Age Children

Table 6-4 shows the difference in the number of poor school-age children from the 1990 census for categories of counties for each of the four candidate models and two of the procedures that rely more heavily on the 1980 census—the stable shares procedure (i) and the averaging procedure (iv). The measure shown is the algebraic difference by category, which is the sum for all counties in a category of the algebraic (signed) difference between the model estimate of poor school-age children and the 1990 census estimate for each county, divided by the sum of the census estimates for all counties.[13] Counties are grouped into five or six categories for each of 11 characteristics—those that were considered in the assessment of the county model regression output discussed above.[14]

The measure in Table 6-4 expresses model-census differences for groups of counties in terms of numbers of poor children, similar to the overall average absolute difference in column 1 of Table 6-3. However, the category difference is expressed as an algebraic measure in which positive differences (overpredictions) within a category offset negative differences (underpredictions). The measure is intended to identify instances of potential bias in a model's predictions. For example, the model may over(under)predict, on average, the number of poor school-age children in larger counties relative to smaller counties.

If the census estimates are a reasonably accurate standard for comparison, sizable category differences between model and census estimates would be disturbing. They would indicate that the errors in the model estimates are not random errors (which occur in any set of estimates), but occur in part because the model systematically over(under)predicts poverty in certain types of counties. Indeed, bias, in terms of over(under)prediction for different types of counties, is arguably more important than the overall absolute difference in evaluating a model that is used repeatedly because there is the risk that the bias will operate for the same areas on each occasion.[15] Although one would not want to use a model that had a large overall absolute difference from the standard of comparison, a model that performed somewhat worse in overall terms but exhibited fewer and less severe biases than another model would be preferable to it.

[13]The formula for counties (*i*) in each category (*j*) is

$$\Sigma_i \left(Y_{\text{model } ij} - Y_{\text{census } ij}\right) / \Sigma_i Y_{\text{census } ij} \,.$$

[14]In addition to the algebraic difference for each category for the four candidate models and four procedures, Appendix C shows for each of them the average proportional algebraic difference; that is, the category difference expressed in terms of percentage errors for counties instead of numbers of poor children (see Tables C-1 and C-2). Differences between the two measures can help identify particular types of counties within a category for which a model performs less well than others.

[15]A search for potential biases is also important to identify possible approaches to model improvement.

TABLE 6-4 Comparison of Model Estimates with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Algebraic Difference by Category of County (in percent)

	Model				Other Procedures		
Category	Log Number Under 21 (a)	Log Number Under 18 (b)	Log Rate Under 21 (c)	Log Rate Under 18 (d)	Stable Shares (i)	Average of Census and (a) (iv)	Number of Counties[a]
Census Division[b]							
New England	–2.9	–2.9	–2.9	–2.9	35.9	7.8	67
Middle Atlantic	–2.8	–2.8	–2.8	–2.8	27.1	4.4	150
East North Central	–0.2	–0.2	–0.2	–0.2	–2.8	–5.6	437
West North Central	1.7	1.7	1.7	1.7	–1.8	–2.1	618
South Atlantic	0.5	0.5	0.5	0.5	14.8	8.1	591
East South Central	–4.5	–4.5	–4.5	–4.5	14.1	2.1	364
West South Central	–2.7	–2.7	–2.7	–2.7	–18.1	–6.3	470
Mountain	4.3	4.3	4.3	4.3	–23.2	–3.1	281
Pacific	6.5	6.5	6.5	6.5	–21.3	0.2	163
Metropolitan Status							
Central county of metropolitan area	2.4	1.6	–0.1	–0.5	–1.6	0.4	493
Other metropolitan	–6.6	–5.0	5.1	6.3	3.2	3.4	254
Nonmetropolitan	–4.2	–2.8	–0.3	0.4	3.3	–1.4	2394
1990 Population Size							
under 7,500	–9.0	–2.3	–1.9	2.3	16.5	1.3	525
7,500-14,999	–4.4	0.5	2.5	5.5	10.9	2.2	630
15,000-24,999	–5.1	–2.6	0.3	1.9	6.2	–0.6	524
25,000-49,999	–4.2	–2.9	0.6	1.3	2.4	–1.3	620
50,000-99,999	–3.5	–5.1	–1.2	–2.3	–2.5	–3.3	384
100,000-249,999	–1.8	–4.4	–1.8	–3.5	–4.9	–3.3	259
250,000 or more	3.3	3.2	0.5	0.5	–0.6	1.8	199

continued on next page

TABLE 6-4 Continued

	Model				Other Procedures		
Category	Log Number Under 21 (a)	Log Number Under 18 (b)	Log Rate Under 21 (c)	Log Rate Under 18 (d)	Stable Shares (i)	Average of Census and (a) (iv)	Number of Counties[a]
1980 to 1990 Population Growth							
Decrease of more than 10.0%	–1.9	0.6	–3.4	–1.9	9.1	–3.4	444
Decrease of 0.1-10.0%	–0.6	–0.5	–1.9	–1.8	7.5	–2.7	972
0.0-4.9%	–2.8	–2.8	–3.2	–3.1	11.0	–0.2	547
5.0-14.9%	0.0	–1.0	0.2	–0.6	6.1	2.1	620
15.0-24.9%	7.7	5.8	5.5	4.6	–12.8	2.4	260
25.0% or more	–4.0	–1.4	1.7	3.1	–21.2	1.0	292
Percent Poor School-Age Children, 1980							
Less than 9.4%	–4.0	–4.5	0.0	0.2	2.4	–1.1	516
9.4-11.6%	–0.5	–1.0	–1.6	–1.8	–9.9	–3.6	524
11.7-14.1%	3.6	2.3	1.8	1.0	–4.2	0.2	530
14.2-17.2%	0.9	1.2	–1.2	–1.4	–5.0	–1.8	523
17.3-22.3%	1.8	1.7	0.3	–0.1	10.7	4.2	519
22.4-53.0%	–2.2	0.8	1.3	2.8	12.3	4.1	523
Percent Hispanic, 1990							
0.0-0.9%	–3.4	–3.3	–1.6	–1.5	10.7	0.2	1770
1.0-4.9%	0.5	0.1	0.4	0.1	0.2	–0.4	847
5.0-9.9%	–1.4	–0.6	–1.1	–0.8	6.7	1.7	193
10.0-24.9%	2.2	1.8	0.7	0.5	–5.7	0.1	181
25.0-98.0%	3.9	4.6	2.2	2.7	–16.8	–0.4	150

Percent Black, 1990							
0.0-0.9%	–1.2	0.3	3.9	4.9	–3.7	–0.5	1446
1.0-4.9%	–0.7	–2.0	1.3	0.5	–6.3	–2.9	615
5.0-9.9%	–2.9	–2.5	–0.7	–0.6	–8.4	–1.8	294
10.0-24.9%	2.0	1.2	–1.0	–1.3	–2.6	0.2	381
25.0-87.0%	1.0	1.7	–1.8	–1.4	16.5	3.7	405
Persistent Rural Poverty, 1960-1990[c]							
Rural, not poor	–4.0	–3.7	–1.2	–1.0	0.1	–3.4	1740
Rural, poor	–5.0	–2.1	0.7	2.1	9.8	1.2	535
Not classified	1.7	1.2	0.3	0.0	–1.2	0.7	866
Economic Type, Rural Counties[c]							
Farming	–5.5	–2.5	–1.6	0.7	13.2	1.1	556
Mining	–10.7	–5.1	–6.3	–3.6	–8.9	–10.6	146
Manufacturing	–6.2	–5.9	–1.7	–1.0	12.1	–0.2	506
Government	2.1	–1.3	6.3	3.2	–0.9	0.0	243
Services	–3.9	–3.0	–1.8	–1.2	–5.8	–4.3	323
Nonspecialized	–3.7	–1.0	–0.1	1.4	2.2	–1.5	484
Not classified	1.7	1.2	0.3	0.0	–1.2	0.7	883
Percent Group Quarters Residents, 1990							
Less than 1.0%	–6.7	–2.7	2.0	4.7	–1.4	0.3	545
1.0-4.9%	0.3	0.7	–0.3	0.1	–0.4	0.1	2187
5.0-9.9%	2.3	–4.4	0.5	–5.2	7.8	–0.8	299
10.0-41.0%	14.2	–3.2	7.4	–7.5	1.8	–2.2	110

continued on next page

TABLE 6-4 Continued

Category	Model				Other Procedures		
	Log Number Under 21 (a)	Log Number Under 18 (b)	Log Rate Under 21 (c)	Log Rate Under 18 (d)	Stable Shares (i)	Average of Census and (a) (iv)	Number of Counties[a]
Status in CPS, 1989-1991							
In CPS sample	1.4	1.0	–0.2	–0.5	–0.6	0.5	1028
In CPS, no poor children aged 5-17	–2.6	–1.9	7.3	7.8	10.0	5.9	246
Not in CPS sample	–4.1	–2.8	–0.1	0.6	0.6	–2.3	1867
Change in Poverty Rate for School-Age Children, 1980-1990							
Decrease of more than 3.0%	7.5	10.4	16.2	18.1	51.6	30.0	536
Decrease of 0.1-3.0%	2.1	1.9	3.1	2.9	29.2	12.1	649
0.0-0.9%	–2.6	–0.8	–0.4	0.5	4.3	3.1	272
1.0-3.4%	3.8	2.2	3.4	2.6	–5.1	0.2	621
3.5-6.4%	–1.2	–2.4	–3.8	–4.3	–14.3	–8.3	532
6.5-38.0%	–7.2	–5.2	–8.7	–7.8	–25.2	–14.5	523

NOTES: The census estimates are controlled to the CPS national estimate for 1989. The algebraic difference by category is the sum for all counties in a category of the algebraic (signed) difference between the model estimate of poor school-age children and the 1990 census estimate for each county, divided by the sum of the census estimates for all counties in the category. See text for definitions of models.

[a]3,141 counties are assigned to a category for most characteristics; 3,135 counties are assigned to a category for 1980-1990 population growth and 1980 percent poor school-age children; 3,133 counties are assigned to a category for 1980-1990 percent change in poverty rate for school-age children.

[b]Census division states:

New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
Middle Atlantic: New York, New Jersey, Pennsylvania
East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin
West North Central: Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, Kansas
South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
East South Central: Kentucky, Tennessee, Alabama, Mississippi
West South Central: Arkansas, Louisiana, Oklahoma, Texas
Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada
Pacific: Washington, Oregon, California, Alaska, Hawaii

[c]The Economic Research Service, U.S. Department of Agriculture, classifies rural counties by 1960-1990 poverty status and economic type. Counties not classified are urban counties and rural counties for which a classification could not be made.

SOURCE: Data from U.S. Census Bureau.

The panel drew several general conclusions from Table 6-4 about the performance of alternative county models in predicting numbers of poor school-age children for categories of counties:

- The performance of the four candidate models is similar. However, the log number (under 18) model (b) performs somewhat better than the log rate (under 21) model (c), which in turn performs better than the other two, the log number (under 21) model (a) and the log rate (under 18) model (d).

Performance in this instance is evaluated principally in terms of the spread among the differences for categories of counties (the spread between the largest positive and negative category differences for a characteristic). A better performing model has a narrower spread for a greater number of characteristics than other models. As an example (see Table 6-4), the spread among the category differences for counties classified by percentage of group quarters residents is 5.1 percentage points for model (b), 7.7 percentage points for model (c), 12.2 percentage points for model (d), and 20.9 percentage points for model (a).

Also entering into the panel's judgment is consideration of the magnitude and pattern of differences: a better performing model has smaller differences from the census and exhibits fewer obvious patterns across categories than other models. Continuing with the same example from Table 6-4, there is no pattern to the category differences for counties classified by percentage of group quarters residents for model (b), whereas model (a) exhibits a strong monotonic pattern in which the number of poor school-age children is overpredicted for counties with higher percentages of group quarters residents relative to counties with lower percentages. Also, the magnitude of the category differences for counties classified by percentage of group quarters residents is small for model (b)—no difference is larger than 5 percent in either direction. In contrast, the category differences for model (a) are as high as 14 percent for one of the categories.

- There are characteristics for which some or all models exhibit poor performance in terms of the spread between the largest and smallest category differences, the pattern of the differences across categories, or the magnitude of the differences (see below, "Category Differences for Specific Characteristics"). There are also some characteristics for which all four models perform well: percentage of poor school-age children in 1980; percentage of black population in 1990; and whether a rural county was persistently poor from 1960 to 1990.
- The four candidate models perform better on most characteristics than the four procedures that rely more heavily on the 1980 census. This is generally true, as discussed below, even for characteristics on which the candidate models perform poorly. However, the averaging procedure (iv), which averages 1980 census estimates and estimates from model (a), performs reasonably well for many characteristics. In contrast, the stable shares procedure (i), which simply ratio adjusts the 1980 census estimates to the CPS national estimate for 1989, performs

substantially worse than all of the models and other procedures on almost every characteristic.

Category Differences for Specific Characteristics

Category differences from the 1990 census estimates are discussed below for characteristics for which Table 6-4 shows that some or all four candidate models exhibit poor performance in comparison with the census in estimating the number of poor school-age children: percentage change from 1980 to 1990 in the poverty rate for school-age children; population growth from 1980 to 1990; 1990 population size; percentage of Hispanic population in 1990; percentage of group quarters residents in 1990; and census geographic division.

Percentage Change from 1980 to 1990 in Poverty Rate for School-Age Children All four candidate models show a pronounced pattern of overpredicting the number of poor school-age children in counties that experienced the greatest decline in the poverty rate for school-age children from 1980 to 1990 and, conversely, underpredicting the number of poor school-age children in counties that experienced the greatest increase in the poverty rate for school-age children in that period. The category differences are smaller for the log number models (a, b) than for the log rate models (c, d): the spread between the largest positive and largest negative differences is 15-16 percentage points for models (a) and (b) and 25-26 percentage points for models (c) and (d).

One would not expect any of the candidate models to perform particularly well in predicting the number of poor school-age children for the counties at the extremes of the distribution of change in the poverty rate from 1980 to 1990. This variable is closely related to the variable that the models are trying to estimate, and the process of fitting a regression line to all of the data will generally not result in good predictions for the extreme values of the distribution. In other words, one would expect the models to perform less well for counties that experienced the largest changes (increase or decrease) in the poverty rate for school-age children.

Despite the large differences for some categories of this characteristic, however, the four candidate models perform substantially better than the procedures that rely more heavily on the 1980 census—see Table 6-4. (See also Figure 6-1, which shows the category differences for percentage change in the school-age poverty rate from 1980 to 1990 for the log number (under 21) model (a), the log number (under 18) model (b), the stable shares procedure (i), and the averaging procedure (iv).) The stable shares procedure performs very poorly: because it assumes the same proportional distribution of poor school-age children in 1989 as in 1979 (from the 1980 census), by definition it will miss any change in poverty rates that occurred over time. The procedure (iv) that averages the estimates from the 1980 census and the log number model (under 21) for 1989

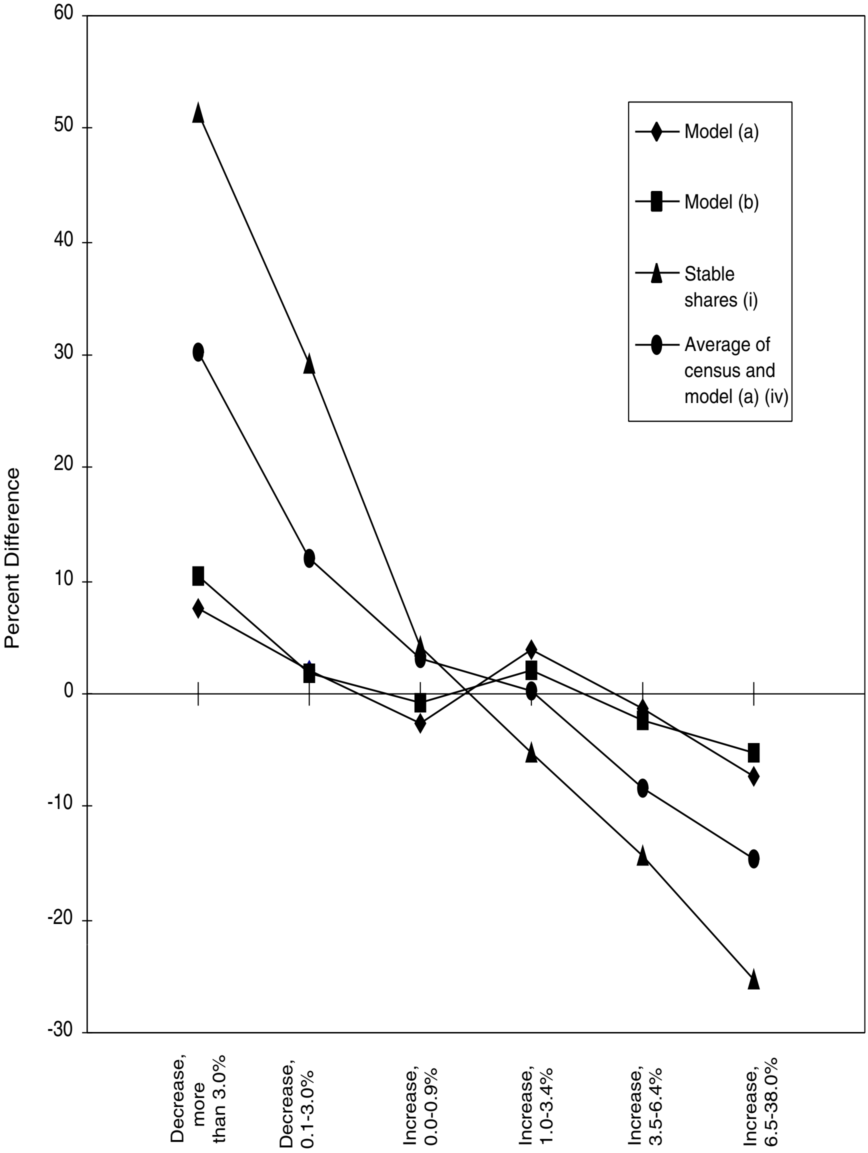


FIGURE 6-1 Change in poverty rate for school-age children, 1980-1990: Category differences from the 1990 census.

performs better than the stable shares procedure but not nearly as well as the four candidate models (two not shown in Figure 6-1).

Population Growth from 1980 to 1990 All four candidate models tend to overpredict the number of poor school-age children in counties that experienced larger population increases from 1980 to 1990 relative to counties that experienced smaller increases or declines in population. The exception to a generally monotonic pattern is that the four models underpredict the number of poor school-age children for counties that experienced population increases of 25 percent or more relative to counties that experienced increases of 15-25 percent. The log number (under 21) model (a) has the largest spread in category differences for this characteristic of the four candidate models—12 percentage points between the largest positive and negative differences.

The stable shares estimation procedure (i) performs very poorly on this characteristic. In contrast to the four candidate models, it overpredicts the number of poor school-age children in counties that experienced declines or smaller increases in population from 1980 to 1990 relative to counties that experienced larger population increases. The spread between the largest positive and negative category differences for the stable shares procedure is 32 percentage points. The averaging procedure (iv) exhibits small differences for population growth categories (see Figure 6-2).

1990 Population Size The four candidate models vary in their performance for counties classified by population size. The log number (under 21) model (a) tends to overpredict the number of poor school-age children in larger size counties relative to smaller size counties. The log number (under 18) model (b) and the log rate (under 21) model (c) do not show a particular pattern to the category differences for this characteristic, and the category differences are not large. The four candidate models perform better than the stable shares model (i), which relies solely on 1980 census data. However, the model (iv) that averages 1980 census estimates with estimates from the log number (under 21) model (a) for 1989 performs reasonably well in predicting numbers of poor school-age children for county population size categories (see Figure 6-3).

Percentage of Hispanic Population in 1990 All four candidate models tend to overpredict the number of poor school-age children in counties with larger percentages of Hispanics relative to counties with smaller percentages, but the spread between the largest positive and negative differences is small. When the category differences are measured in proportionate terms for counties instead of in terms of numbers of poor school-age children, the models tend to *underpredict* the number of poor school-age children in counties with larger percentages of Hispanics (see Appendix C). The different patterns of the two category differ-

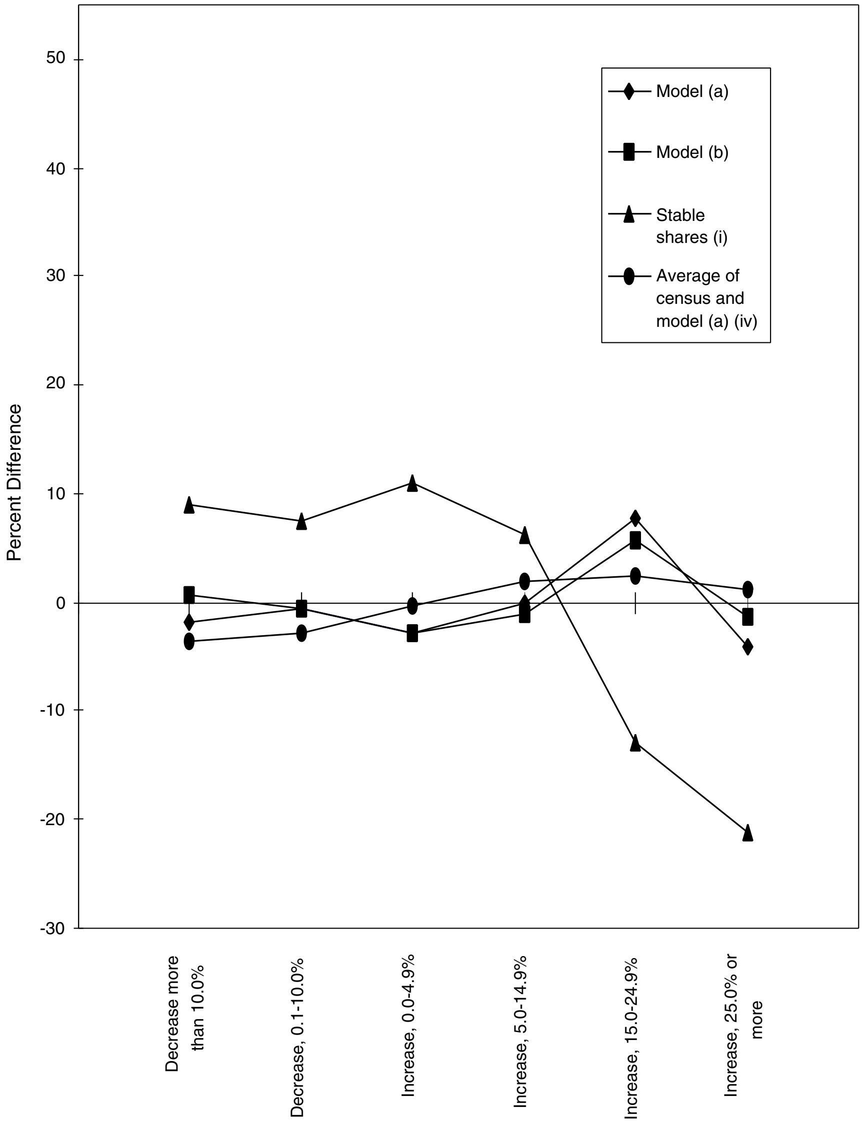


FIGURE 6-2 Population growth, 1980-1990: Category differences from the 1990 census.

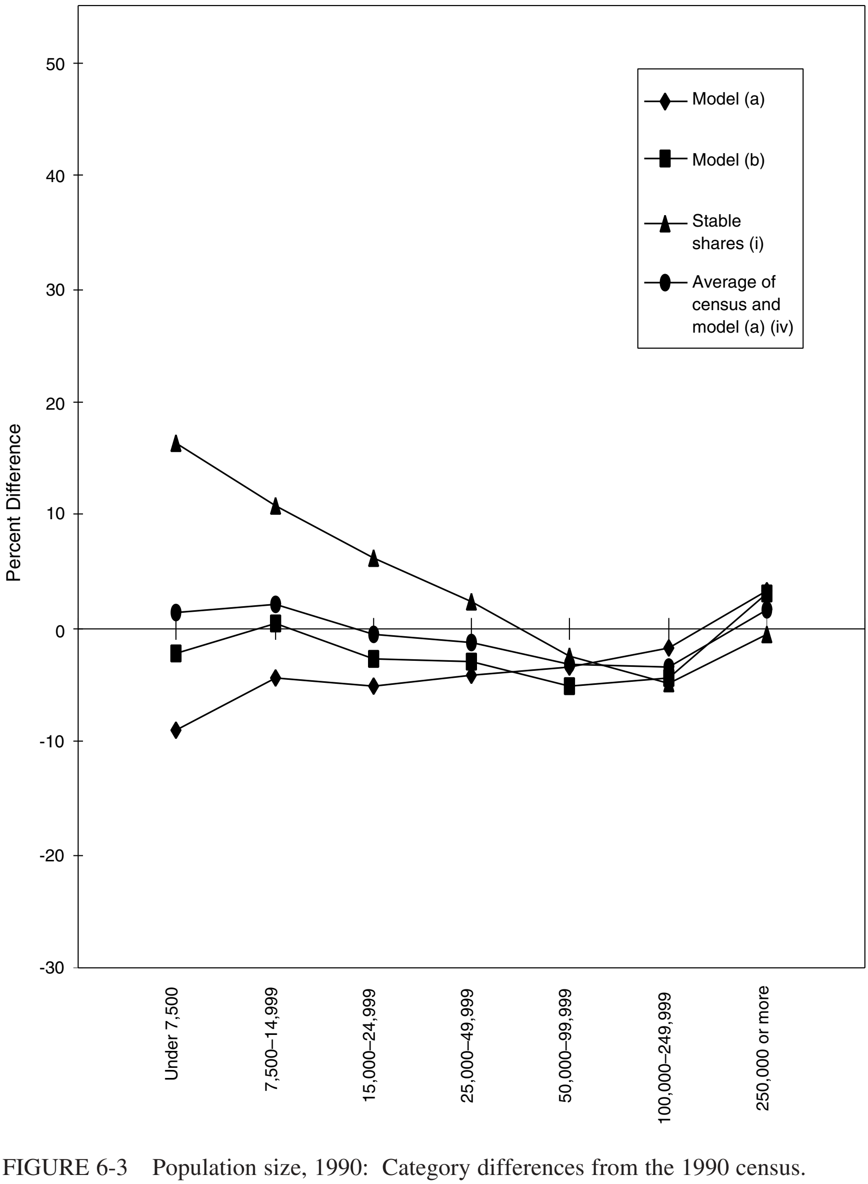


FIGURE 6-3 Population size, 1990: Category differences from the 1990 census.

ence measures suggest that the models may perform differently for small counties with many Hispanics (primarily rural border counties) and large counties (cities).

The stable shares procedure (i), which relies solely on the 1980 census estimates, performs poorly on this characteristic. However, the averaging procedure (iv) performs reasonably well (see Figure 6-4).

Percentage of Group Quarters Residents in 1990 The four candidate models vary in their performance for counties classified by percentage of group quarters residents. The log number (under 21) model (a) substantially overpredicts the number of poor school-age children in counties with larger proportions of group quarters residents relative to other counties. The log rate (under 21) model (c) shows a similar but less pronounced pattern of category differences. The log rate (under 18) model (d) shows the opposite pattern, in which it underpredicts the number of poor school-age children in counties with larger proportions of group quarters residents relative to other counties. In contrast, the category differences for the log number (under 18) model (b) are small and do not show a pronounced pattern across categories of this characteristic.

When the evident bias in predicting the number of poor school-age children in counties relative to their percentage of group quarters residents was discovered in the first round of evaluations of model (a), the Census Bureau developed model (b) to ameliorate the problem, with the desired result. The reasoning was as follows. In model (a), the two predictor variables—total child exemptions (assumed to be under age 21) from IRS tax records and the population estimate of the under 21 age group—are used together to estimate the number of people under age 21 in families that do not file tax returns. These families are assumed to be poorer, on average, than families that file tax returns. As can be seen from Tables 6-1 and 6-2, the regression coefficients for these two variables are of similar magnitude but of opposite sign.

However, in counties with large percentages of group quarters residents under age 21, primarily college students and military personnel, the relationship between the IRS variable and the population estimate may be distorted. To the extent that college students and military personnel under age 21 live in a county that is not the same as the county in which their parents reside or file tax returns, they will not be recorded as child exemptions in their county of residence. Consequently, there will be an overestimate of the number of people under age 21 in families that do not file returns in these counties and a corresponding overestimate, through the model, of the number of school-age children in poverty.

Model (b) replaces the population estimate for the under 21 age group as a predictor variable with the population estimate for the under 18 age group. This change not only eliminates the pattern of overpredicting the number of poor school-age children as a function of the percentage of group quarters residents that is so pronounced in model (a), but it also causes model (b) to perform better than model (a) on a number of other characteristics (e.g., population size). For

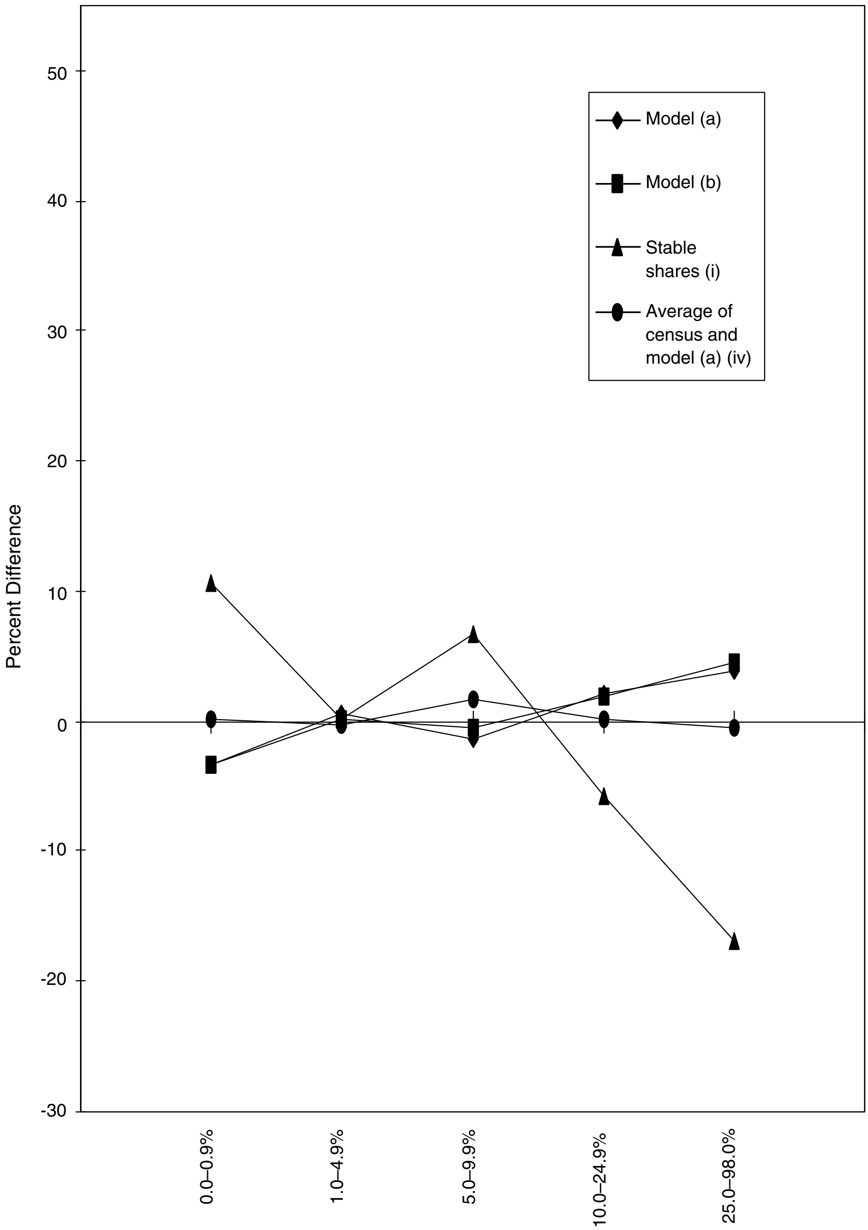


FIGURE 6-4 Percent Hispanic population, 1990: Category differences from the 1990 census.

reasons that are not clear, the under 18 formulation does not improve the performance of the log rate model; in fact, the log rate (under 18) model (d) generally performs worse than the log rate (under 21) model (c).

Interestingly, the procedures that rely more heavily on the 1980 census (i-iv)—even the stable shares procedure—perform reasonably well in predicting the number of poor school-age children for counties categorized by percentage of group quarters residents (see Figure 6-5).

Census Division All four candidate models show differences from the census for counties categorized by census division. In particular, the four models overpredict the number of poor school-age children in counties in the West (in the Mountain Division and, particularly, in the Pacific Division) relative to counties in other areas. The spread between the largest positive and negative differences is 11 percentage points.

Because the county estimates from the four candidate models are raked to the state estimates from the Census Bureau's state model and census divisions are combinations of states, category differences on this characteristic must be attributable to the state model.[16] As discussed later, the category differences by area in the state model occurred also in several other years and warrant further investigation (see below, "State Model"). Yet the state raking procedure, which is done for the four candidate models and for the procedures that assume stable shares within state and stable rates within state (ii, iii), results in substantially better performance on this characteristic than the stable shares procedure (i). The averaging procedure (iv), which partly reflects the effects of the state raking, also performs better than the stable shares procedure (see Figure 6-6).

Differences in Proportions of Poor School-Age Children

The panel examined category differences in estimates of proportions (rather than numbers) of poor school-age children in a form similar to Table 6-4 and reached the same conclusions. Comparisons were performed only for the four candidate models, not for the other procedures.

First, the performance of the four candidate models is similar. Second, the two models that performed best in estimating the number of poor school-age children—log number (under 18) model (b) and log rate (under 21) model (c)—also perform best in estimating the proportion of poor school-age children. However, model (c) performs slightly better than model (b) in estimating proportions,

[16]The category differences are the same for all four candidate models because they are raked to the same set of state estimates (see Table 6-4). The average proportional category differences shown in Appendix C vary somewhat because they are calculated relative to each county's 1990 census estimated number of poor school-age children before being summed (see Table C-2).

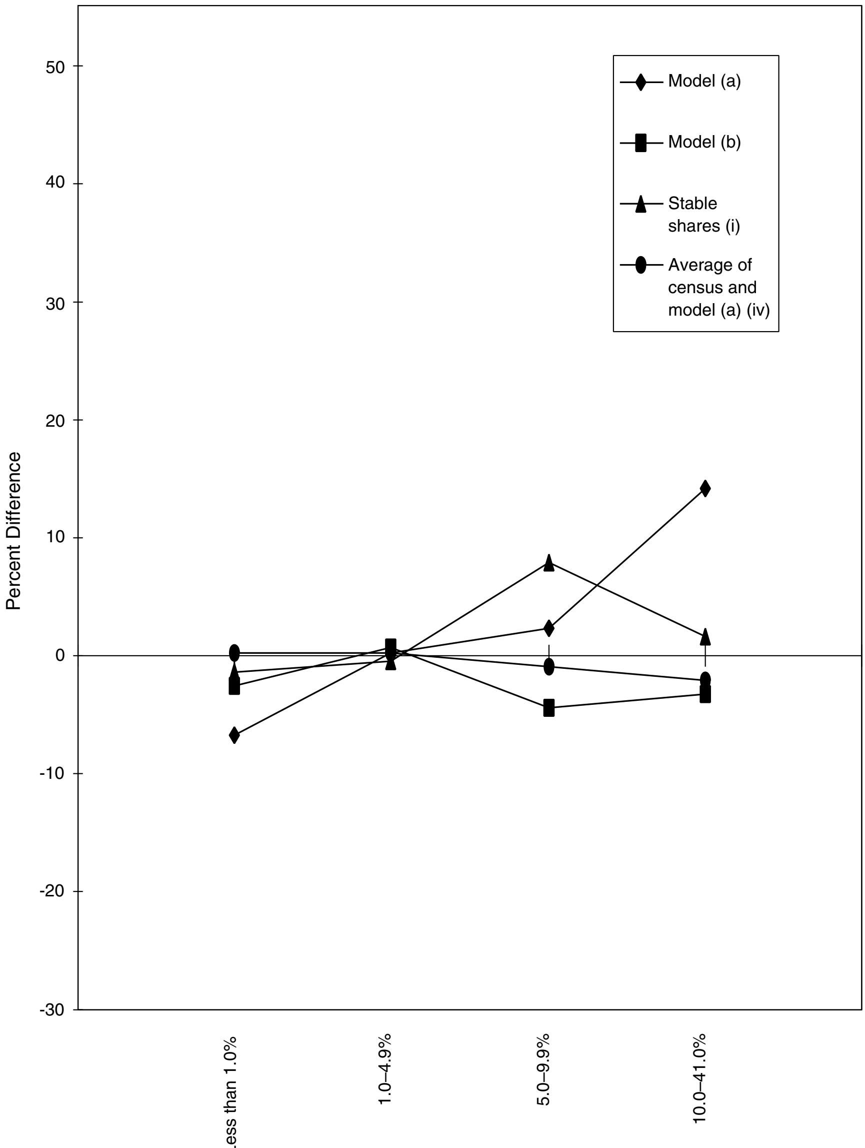


FIGURE 6-5 Percent group quarters residents, 1990: Category differences from the 1990 census.

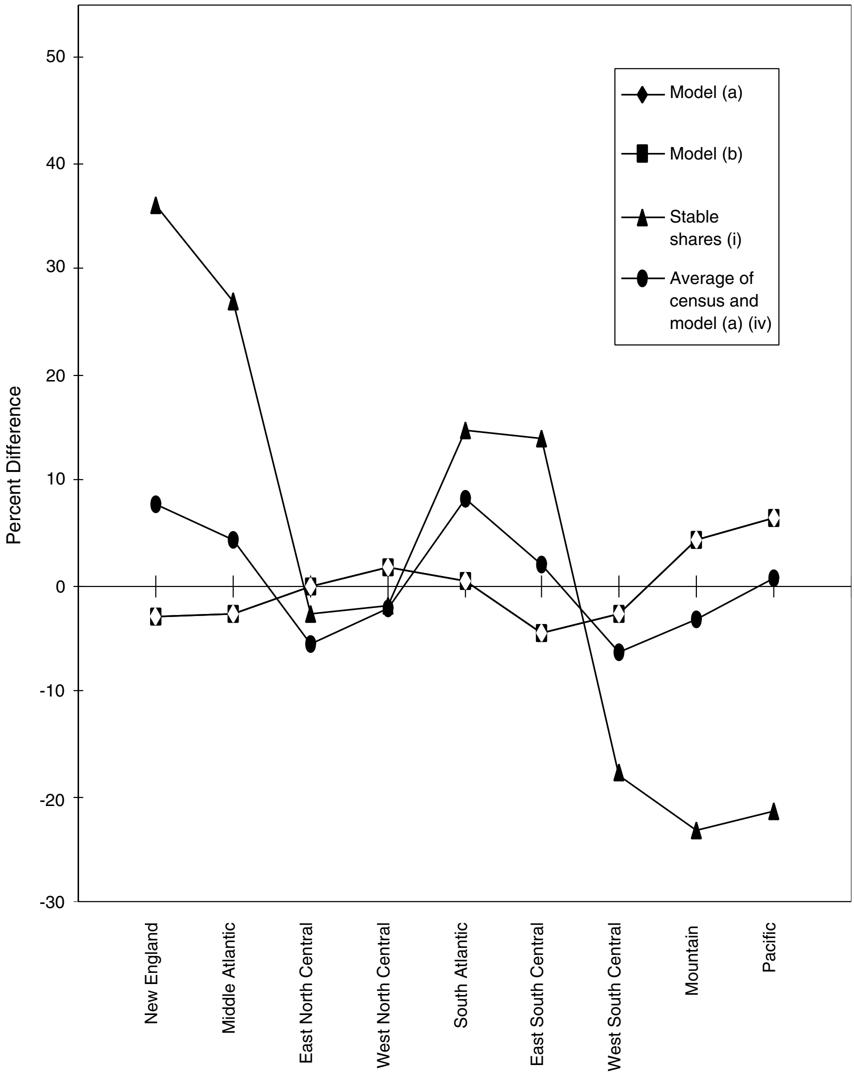


FIGURE 6-6 Census division: Category differences from the 1990 census.

while model (b) performs slightly better than model (c) in estimating numbers of poor school-age children. This reversal is expected because the use of population estimates for children aged 5-17, which themselves contain errors, to convert estimated numbers to estimated proportions from the log number models puts these models at a disadvantage for comparisons of proportions. Conversely, the use of population estimates for children aged 5-17 to convert estimated proportions to estimated numbers from the log rate models puts these models at a disadvantage for comparisons of numbers (see Chapter 8).

Poverty rates (proportions poor) of school-age children enter the Title I allocation formulas as thresholds, so the panel and the Census Bureau examined the correspondence between each of the four candidate models and the 1990 census in classifying counties and school-age children into three poverty rate categories: 0 to 15 percent; 15 to 30 percent; and 30 percent or higher. (See Table 6-5; no comparisons were performed for the other procedures.) A poverty rate of 15 percent or higher is an eligibility threshold for concentration grants; 15 percent and 30 percent poverty rates are thresholds for hold-harmless provisions of the allocation formulas.

When there are two poverty rate categories, 0 to 15 percent and 15 percent or higher, each of the four candidate models performs equally well, assigning about 87 percent of the counties, which include about 92 percent of the poor school-age children, to the same category as the 1990 census (column 5, top half and bottom half of Table 6-5). When there are three poverty rate categories, 0 to 15 percent, 15 to 30 percent, and 30 percent or higher, each of the four candidate models assigns about 81 percent of the counties, which include about 88 percent of the poor school-age children, to the same category as the 1990 census (column 6, top half and bottom half of Table 6-5).

CPS-Census Differences

A possible explanation of some of the category differences identified in the 1990 census comparisons just described may be, not that a model is in error, but that measurement of poverty differs systematically between the census and the CPS because of the many differences in data collection procedures (see Chapter 3). The Census Bureau performed chi-square tests to determine if there were significant differences between estimates from the March 1990 CPS and the 1990 census of the number of school-age children and the number and proportion poor in this age group in 1989 for county groupings (Fay, 1997).[17] More specifically, the tests determined if the ratios of the CPS and census estimates for categories of a characteristic, such as county population size, were significantly different from each other. The characteristics tested were those examined in the 1990 census comparisons.

The tests generally show inconclusive results. However, there is some evidence that, when compared with the 1990 census, the March 1990 CPS estimates

[17]The March 1990 CPS estimates for the categories involved are direct estimates produced using the CPS weights.

TABLE 6-5 Agreement Between Model Estimates for 1989 and 1990 Census County Estimates for Proportions of School-Age Children in Poverty in 1989 (in percent)

	Counties						
	Model and Census Estimate in Same Poverty Rate Category				Percent in Agreement		
	Under 15%	15% or More	15-30%	30% or More	Under 15% and 15% or More	Under 15%, 15-30%, and 30% or More	
Model		(1)	(2)	(3)	(4)	(5)	(6)
(a) Log Number (Under 21)	30.1	57.0	39.2	11.4	87.1	80.7	
(b) Log Number (Under 18)	30.5	57.1	38.4	11.9	87.6	80.8	
(c) Log Rate (Under 21)	28.8	58.6	40.1	12.9	87.4	81.8	
(d) Log Rate (Under 18)	28.4	58.6	39.5	13.0	87.0	80.9	

	School-Age Children					
	Model and Census Estimate in Same Poverty Rate Category				Percent in Agreement	
Model	Under 15%	15% or More	15-30%	30% or More	Under 15% and 15% or More	Under 30%, 15-30%, and 30% or More
(a) Log Number (Under 21)	40.7	51.0	39.9	7.3	91.7	87.9
(b) Log Number (Under 18)	40.9	50.7	38.5	7.3	91.6	86.7
(c) Log Rate (Under 21)	40.1	51.5	41.0	7.6	91.6	88.7
(d) Log Rate (Under 18)	40.3	51.1	40.3	7.5	91.4	88.1

NOTE: Census estimates are controlled to the CPS national estimate for 1989.
SOURCE: Data from U.S. Census Bureau.

higher numbers and proportions of poor school-age children in metropolitan counties and larger-size counties relative to medium-size counties. (CPS estimates for small-size counties have low reliability because of the relatively small proportion of the population in such counties and the small number of these counties in the CPS sample.) Also, while not significant, a pattern is evident in which the March CPS, when compared with the 1990 census, tends to estimate higher numbers and proportions of poor school-age children in counties with higher percentages of Hispanic population. These results for population size and percentage of Hispanic population parallel the results from the 1990 census comparisons described above. They suggest that at least some portions of the category differences for the candidate models for these two characteristics arise from differences in the CPS measurement of poverty and are not due to model error as such. Whether similar CPS-census differences would be present for 1993 or 1995 is, of course, not known.

Summary

Keeping in mind the limitations of a single census-based validation opportunity, the panel concluded that the four candidate models perform substantially better in predicting the number and proportion of poor school-age children for counties for 1989 than the simple stable shares procedure (i), which relies solely on estimates from the previous (1980) census and the current (1989) CPS national total. Using the state model to rake the 1980 census county estimates for consistency with updated estimates of poor school-age children in each state, as is done in procedures (ii) and (iii), is an improvement over procedure (i). However, the four candidate models, which use a county regression model together with the state model, perform much better than procedures (ii) and (iii). Finally, the four candidate models perform better in many respects than procedure (iv), which averages the 1980 census estimates and the 1989 estimates from the log number (under 21) model (a), although this averaging procedure shows good performance on some characteristics. Overall, the comparisons with the procedures that rely more heavily on the 1980 census provide significant evidence in favor of a model-based approach for updated estimates of poor school-age children and against using estimates that derive solely or largely from the previous census.

The panel further concluded that, while the performance of the four candidate models in comparison with the 1990 census is broadly similar, when consideration is given to measures of overall absolute difference and differences for categories of counties, for estimates of numbers and estimates of proportions of poor school-age children, the log number (under 18) model (b) and the log rate (under 21) model (c) perform better than the other two. Comparing models (b) and (c), model (b) performs somewhat better, and the Census Bureau used this model to prepare the revised county estimates of poor school-age children in 1993. The comparisons also identify areas of performance of model (b) that deserve further examination in an ongoing research program to continue to improve model-based estimates of poverty for small geographic areas.

Comparisons with the CPS

For the 1995 county model external evaluations, the emphasis shifted to finding a way to look for persistent bias. An apparent bias identified in a single validation, such as the 1990 census comparisons summarized above, may be a one-time effect that will not occur in other years for which a model is estimated. For any particular year, it is almost inevitable that the differences between the model estimates and target values will be somewhat larger for some categories of counties than others. But if such differences persist for the same categories of counties over time, some areas may continually receive more funding and other areas may continually receive less funding than if the true values were known.

As a type of external validation by which the issue of persistent bias could be examined, the panel and the Census Bureau compared estimates of poor school-age children from the 1995 county model for categories of counties for 1989, 1993, and 1995, with CPS direct estimates for those categories for the three periods. Three years of CPS data were used to form the weighted estimates in each case in order to reduce the sampling variability.[18]

Table 6-6 shows the difference in the number of poor school-age children from the county model, estimated for 1989 (using corrected IRS data), 1993, and 1995, and the weighted 3-year CPS direct estimates centered on those years for categories of counties. The measure shown is the algebraic difference by category, which is the sum for all counties in a category of the algebraic (signed) difference between the model estimate of poor school-age children and the weighted CPS direct estimate, divided by the sum of the weighted CPS direct estimates for the category.

Comparisons with weighted CPS direct estimates have the advantage over comparisons with the census that they can be performed for multiple years. They have the disadvantage that the sample sizes for CPS estimates, even aggregated for 3 years, are small for many categories of counties, thus making the comparisons much more uncertain than the 1990 census comparisons because of the much greater variability in the standard of comparison. Also, in analyzing the CPS comparisons, one must keep in mind that the model estimates are raked to the state estimates, which are developed from a single year of the CPS.

The model-CPS aggregate differences in Table 6-6 differ widely among

[18]This analysis is not the same as the analysis of regression output described above, in which the standardized residuals from the model for counties with sampled households in the CPS—representing the standardized differences between the model estimates and the direct estimates on the log scale—were examined for categories of counties.

TABLE 6-6 Comparison of County Model Estimates with CPS Aggregate Estimates of the Number of Poor School-Age Children, 1995, 1993, and 1989: Algebraic Difference by Category of County (in percent)

Category	No. of Counties[a] (1)	Model-CPS, 1995[b] (2)	Model-CPS, 1993[b] (3)	Model-CPS, 1989[b] (4)	Sample Size, CPS 1996[c] (5)
Census Region[d]					
Northeast	217	–2.87	0.81	–4.36	10,708
Midwest	1,055	–0.49	0.61	–4.31	11,393
South	1,425	4.05	–0.13	4.48	15,440
West	444	–4.16	–0.95	–0.43	12,141
Census Division[d]					
New England	67	–13.51	1.87	27.07	3,696
Middle Atlantic	150	0.05	0.54	–9.79	7,012
East North Central	437	–6.10	–0.64	–3.04	6,841
West North Central	618	18.31	4.25	–7.44	4,552
South Atlantic	591	1.82	0.83	4.12	8,150
East South Central	364	–5.53	–5.85	9.32	2,529
West South Central	470	12.00	1.90	2.44	4,761
Mountain	281	–3.91	19.87	0.84	5,543
Pacific	163	–4.24	–6.48	–0.92	6,598
Metropolitan Status					
Central county of metropolitan area	493	–2.75	–0.91	–3.53	34,343
Other metropolitan	254	53.75	–3.64	8.44	2,801
Nonmetropolitan	2,394	1.24	3.50	8.32	12,538
1990 Population Size					
Under 7,500	525	–17.21	57.03	0.74	933
7,500-14,999	630	19.82	–23.67	–0.19	1,550
15,000-24,999	524	2.94	6.24	17.02	2,289
25,000-49,999	620	30.46	–0.23	–4.46	4,204
50,000-99,999	384	–2.52	4.99	22.47	5,979
100,000-249,999	259	17.27	12.12	–3.88	8,263
250,000 or more	199	–7.24	–2.49	–3.10	26,464
1980 to 1990 Population Growth					
Decrease of more than 10.0%	444	–2.71	–22.03	–4.29	2,170
Decrease of 0.1-10.0%	972	–4.31	2.44	–1.32	10,655
0.0-4.9%	547	6.04	3.41	3.18	8,015
5.0-14.9%	620	1.12	5.97	4.61	11,590
15.0-24.9%	260	–0.07	–4.11	–10.44	9,305
25.0% or more	292	–0.52	–2.27	10.31	7,947

continued on next page

TABLE 6-6 Continued

Category	No. of Counties[a] (1)	Model-CPS, 1995[b] (2)	Model-CPS, 1993[b] (3)	Model-CPS, 1989[b] (4)	Sample Size, CPS 1996[c] (5)
Percentage of Poor School-Age Children, 1980					
Less than 9.4%	516	2.74	7.22	–1.07	14,980
9.4-11.6%	524	1.39	5.28	4.35	12,291
11.7-14.1%	530	–10.01	–6.49	–6.72	9,837
14.2-17.2%	523	1.28	–5.82	0.44	5,217
17.3-22.3%	519	9.32	17.41	0.23	4,623
22.4-53.0%	523	1.05	–14.81	4.11	2,734
Percentage Hispanic, 1990					
0.0-0.9%	1,770	1.26	–0.75	3.13	12,848
1.0-4.9%	847	9.33	1.45	4.32	16,966
5.0-9.9%	193	–2.81	17.24	6.38	6,999
10.0-24.9%	181	–4.02	–5.14	–8.29	7,236
25.0-98.0%	150	–7.90	–3.29	–5.26	5,633
Percentage Black, 1990					
0.0-0.9%	1,446	8.32	8.02	5.09	10,929
1.0-4.9%	615	7.41	1.04	–1.83	10,630
5.0-9.9%	294	5.41	–2.07	0.95	8,646
10.0-24.9%	381	–4.89	–0.75	3.51	13,437
25.0-87.0%	405	–6.85	–2.82	–6.30	6,040
Persistent Rural Poverty, 1960-1990[e]					
Rural, not poor	1,740	–2.62	1.53	5.47	9,734
Rural, poor	535	22.45	–0.15	14.81	1,698
Not classified	866	–1.28	–0.28	–2.68	38,250
Economic Type, Rural Counties[e]					
Farming	556	–24.56	–29.31	–12.41	1,634
Mining	146	46.97	27.59	40.67	901
Manufacturing	506	–7.10	–3.58	–1.51	2,369
Government	243	120.13	27.59	59.39	1,661
Services	323	–12.18	–12.42	–11.86	2,760
Nonspecialized	484	6.99	18.35	23.89	2,018
Not classified	883	–1.18	–0.20	–2.59	38,339
Percentage of Group Quarters Residents, 1990					
Less than 1.0%	545	3.32	22.03	16.60	3,494
1.0-4.9%	2,187	–1.58	–1.27	–1.84	41,648
5.0-9.9%	299	11.90	–1.22	4.51	3,980
10.0-41.0%	110	49.44	–6.28	17.02	560

TABLE 6-6 Continued

Category	No. of Counties[a] (1)	Model-CPS, 1995[b] (2)	Model-CPS, 1993[b] (3)	Model-CPS, 1989[b] (4)	Sample Size, CPS 1996[c] (5)
Change in Poverty Rate for School-Age Children, 1980-1990					
Decrease of more than 3.0%	536	–3.88	–11.16	–10.04	4,038
Decrease of 0.1-3.0%	649	–4.57	2.63	4.44	12,658
0.0-0.9%	272	2.16	–2.75	9.66	5,102
1.0-3.4%	621	–1.07	0.11	–5.06	14,660
3.5-6.4%	532	9.09	–2.60	–0.66	7,507
6.5-38.0%	523	–1.07	5.17	3.98	5,719

[a]3,141 counties are assigned to a category for most characteristics; 3,135 counties are assigned to a category for 1980-1990 population growth and 1980 percentage of poor school-age children; 3,133 counties are assigned to a category for 1980-1990 percent change in poverty rate for school-age children.

[b]The formula, where there are n counties (i) in category (j), Y_{model} is the estimated number of poor school-age children from the county model, and Y_{CPS} is the estimated number of poor school-age children from a 3-year weighted average of the CPS, is

$$\Sigma_i \, (Y_{\text{model } ij} - Y_{\text{CPS } ij}) \, / \, \Sigma_i Y_{\text{CPS } ij} \, .$$

[c]Number of households (unweighted) in the sample for the March 1996 CPS is shown to give an idea of the relative sample sizes for each category. The 3-year weighted averages are based on 3 years' worth of sample, although some sample cases are the same for 2 years because of the rotational design.

[d]Census region and division states:

Northeast
- New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
- Middle Atlantic: New York, New Jersey, Pennsylvania

Midwest
- East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin
- West North Central: Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, Kansas

South
- South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
- East South Central: Kentucky, Tennessee, Alabama, Mississippi
- West South Central: Arkansas, Louisiana, Oklahoma, Texas

West
- Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada
- Pacific: Washington, Oregon, California, Alaska, Hawaii

[e]The Economic Research Service, U.S. Department of Agriculture, classifies rural counties by 1960-1990 poverty status and economic type. Counties not classified are urban counties and rural counties for which a classification could not be made.

SOURCE: Data from U.S. Census Bureau.

categories of counties, in large part because of the small sample sizes for the CPS estimates, even when aggregated for 3 years. Some of the differences are very large, larger than any of the differences seen in the model-1990 census comparisons above. Generally, the larger model-CPS aggregate differences are for categories of counties with smaller numbers of CPS sample households. For example, the model-CPS aggregate differences often exceed 5 percent for counties grouped into the nine geographic divisions, but they are all less than 5 percent for counties grouped into the four geographic regions.[19]

In addition, the model-CPS aggregate differences for 1989 frequently differ from the model-1990 census differences. This finding is expected, given that the measurement of poverty differs between the census and the CPS because of the many differences in data collection procedures.

Despite the sample size limitations, Table 6-6 can inform an assessment of the performance of the county model if the results are used with caution. Of particular interest are instances in which the model-CPS aggregate differences are both large and in the same direction (plus or minus) for all 3 years for which the county model is estimated. Such findings suggest a possible systematic bias in the model that should be investigated to determine the nature of the bias and what steps could be taken to eliminate or reduce it (e.g., by adding a predictor variable to the model). Several persistent patterns are evident in the model-CPS aggregate differences:

- The model shows a tendency to underpredict the number of poor school-age children in the largest counties, those with 250,000 or more population. This finding is consistent with the results from analyzing the distribution of the standardized residuals from the regression output. The extent of the underprediction is not large, but it appears to be significant given the large number of CPS households in the largest counties.
- The model shows a tendency to underpredict the number of poor school-age children in counties with large percentages of Hispanic residents (10% or more). There is a similar, although less pronounced, tendency for the model to underpredict the number of poor school-age children in counties with large percentages of blacks. It is likely that counties with large percentages of Hispanics or blacks are not homogeneous (e.g., large-percentage black counties include both inner-city and rural areas). Hence, further research is needed to determine whether the underprediction is more or less pronounced for particular subgroups of these counties and, consequently, what steps are appropriate to ameliorate the bias in the model.

---

[19] For future evaluations of this type, the standard errors of the differences should be computed so that significant differences between the model estimates and the CPS 3-year aggregate estimates can be identified.

• The model estimates are consistently very different from the weighted CPS estimates for some categories of rural counties classified by economic type. In particular, the model estimates for rural counties characterized as government are much higher than the corresponding weighted CPS estimates. Although the comparisons by economic type are based on small CPS sample sizes, it seems worthwhile to examine some of these counties to see if a reason for these large differences can be found.

• Finally, the model shows a tendency to underpredict the number of poor school-age children in counties that experienced the largest declines in the poverty rate for school-age children from 1980 to 1990. As was noted above, this finding is consistent with the knowledge that any regression model can only partially predict which cases will have the most extreme values of the outcome variable.

Local Assessment of 1993 County Estimates

The panel performed another type of external evaluation of the original 1993 county estimates of poor school-age children—the use of local knowledge.[20] Using the original 1993 model estimates for all 3,143 counties in the United States, the analysis first sought to identify groups of counties for which the 1993 estimates seemed unusually high or low in relation to prior levels and trends (e.g., from 1980 to 1990) in the number and proportion of poor school-age children and known social and economic trends for these groups of counties. Then, local informants—including staff and members of local councils of government, economic development authorities, welfare agencies, state demographic units, state data centers, and other agencies—were contacted to obtain their assessment of the reasonableness of the implied trends in poverty for school-age children given their knowledge of local socioeconomic conditions.[21]

County Analysis

Changes in the number and proportion of poor school-age children implied by the 1993 estimates were examined for counties categorized by several characteristics, including: population size and metropolitan status; population change; percentage of immigrants; college-dominated counties; reservation and Native American counties; for nonmetropolitan counties, whether predominantly agri-

[20]This evaluation was carried out at the University of Wisconsin-Madison by Dr. Paul Voss, a member of the panel, with the assistance of Richard Gibson and Kathleen Morgen (see Voss, Gibson, and Morgen, 1997).

[21]The discussion refers to "implied" trends because the Census Bureau's county model is not designed to directly estimate change over time.

cultural; and several classifications by geographic location (e.g., state and the regions identified by the U.S. Department of Agriculture).

The analysis identified a number of categories of counties for which further investigation of the reasonableness of the 1993 estimates seemed warranted:

- Large metropolitan central city counties had a high implied percentage change in the number of school-age children in poverty between 1989 and 1993—42 percent. This change declined systematically with decreasing size for metropolitan counties and continued to decline to the most remote, rural nonmetropolitan counties, for which the implied change in the number of school-age children in poverty was –6 percent.
- Counties with higher levels of international immigration had higher implied increases in the number and proportion of poor school-age children.
- Counties with higher percentages of Native Americans had lower implied increases in the number and proportion of poor school-age children. There was no particular pattern for counties with reservations.
- Farm counties had an implied decline in the number and proportion of poor school-age children, while nonfarm metropolitan counties had an implied increase.
- When the country was divided into the 26 regions identified by the U.S. Department of Agriculture, several regions were identified on the extremes of change in the number and proportion of poor school-age children. High implied increases were found in the Northern Metropolitan Belt, the Florida Peninsula, the Southwest, Northern New England, Mohawk New York and Pennsylvania, Lower Great Lakes Industrial, Southern Piedmont, and the Northern Pacific Coast. Small implied increases were found in the Central Corn Belt, the Southern Appalachian Coal Region, the Coastal Plain Cotton Region, the Northern Great Plains, and the Rockies, Mormon, Columbia River Region. The single region with an implied decrease in the number and proportion of poor school-age children was the Mississippi Delta.

Some of these implied changes are apparently related to the general effect of population size, discussed above. However, the findings in this regional analysis, in particular, suggested which states and counties to follow up in discussions with local officials.

Local Input

When counties that share certain characteristics appeared also to share a common pattern of change in the number and proportion of poor school-age children, a variety of individuals with local knowledge were contacted. Initially, 70 individuals associated with state data centers or state data center affiliate units were contacted; they provided a series of responses and referrals to other state and local officials. In addition, 26 states that appeared to have a sizable number of counties that shared a common implied trend in poverty for school-age children were targeted for intensive contact.

The nature of responses varied considerably. In some states, the original 1993 county estimates released by the Census Bureau had not been examined, and there appeared to be little interest in discussing them. In other states, the estimates had been looked at, but the general admonitions about standard errors that accompanied their release had dampened interest in studying them in detail. In contrast, several states had carried out in-depth analyses of the estimates. Of the 26 states targeted for intensive follow up, 8 provided detailed explanations (supported by examples) of trends suggested by the original 1993 county estimates, and 7 more states provided in-depth responses supported by their own analyses.

Almost every state agency contacted expressed specific doubts about the original 1993 estimates for one or more counties—too high here, too low there. In general, however, there was no consensus that the trends implied by the original 1993 county estimates were wrong, even in states for which large numbers of counties experienced apparent declines in the number and proportion of poor school-age children. Of the 26 states, 21 provided explanations as to why the original 1993 estimates appeared to show poverty trends in a specific direction or why the direction of change is too difficult to know. The most common explanations included comments about the size of the county, its rural agricultural nature, the fact that it is a diverse metropolitan county, immigration from abroad, and economic growth or economic decline. Occasionally, reference was made to a military base, an Indian reservation, or a university as an explanation for an apparent trend in poverty for school-age children. In three states, concern was expressed about the role of Food Stamp Program data in the estimation model, as these data were deemed to be unreliable.

In summary, a high level of concern was expressed by individuals with local knowledge about the statistical reliability of the original 1993 county estimates, which is largely due to the Census Bureau's own cautions in this regard, coupled with specific county estimates that seem on the basis of local knowledge to be highly doubtful. These concerns notwithstanding, no categories of counties were identified that experienced apparent trends in the number and proportion of poor school-age children between 1989 and 1993 that were not accepted by local informants. Although the trends for a few counties were not accepted locally, the analysis found no strong indicators of potential bias for groups of counties sharing common characteristics in the county model.

Summary

Considering the external evaluations of alternative models that were conducted by comparison with 1990 census estimates, the external evaluations of 3 years of estimates that were conducted for the 1995 county model by comparison with weighted direct CPS estimates, and the local assessment of the 1993 county estimates, the panel concluded that the county model is working reasonably well. However, further investigation is needed of categories of counties for which the model appears to overpredict or underpredict the number of poor school-age children, particularly when that phenomenon is evident for several periods.

STATE MODEL EVALUATION

The state model plays an important role in the production of county estimates of poor school-age children. Evaluations conducted of the state model for the assessment of the revised 1993 county estimates included an internal evaluation of the regression output for 1989 and 1993 and an external evaluation that compared 1989 estimates from the model with 1990 census estimates of proportions of poor school-age children. The results in each case supported the use of the model. However, the state model evaluations were more limited than the county model evaluations, as alternative state model formulations were not evaluated explicitly.

For the assessment of the 1995 county estimates, further evaluations were conducted of the state model. In particular, the model was estimated for 7 years—1989, 1990, 1991, 1992, 1993, 1995, and 1996—and the regression output for those years was examined to determine if there were any systematic biases in the model estimates. (The model was not estimated for 1994 because the redesign of the CPS sample, consequent to the 1990 census, was partly but not completely phased in for the March 1995 CPS.) Also, there was an evaluation of the state raking factors for 1993 and 1995.

State Model Regression Output

The state regression model is a poverty rate model with the variables not transformed (see equation (2) in Chapter 4). The analysis of the regression output for the state model, estimated for each year from 1989 through 1993 and for 1995 and 1996, examined the same assumptions that were examined for the 1995 county model estimated for 1989, 1993, and 1995. The analysis is somewhat less informative for the state model than for the county model because there are about 1,000 counties with poor school-age children in the CPS, but only 51 states (including the District of Columbia), and states are collectively much more homogeneous than counties with respect to poverty rates and other characteristics. In addition, with respect to both internal and external evaluation, some categories of states do not contain enough states for analysis, thereby reducing the utility of evaluation.

Nonetheless, examination of the regression output for the state model helps assess the validity of its assumptions. With a few exceptions, the analysis supports the assumptions underlying the state model (see below); there is little evidence of significant problems with the model formulation (although there may be other models that fit just as well).

Linearity

Plots of standardized residuals against the four predictor variables in the state model—the proportion of child exemptions reported by families in poverty on tax returns, the proportion of people receiving food stamps, the proportion of people under age 65 who were not included on a tax return, and a residual from the analogous regression equation using the previous census estimate as the dependent variable—support the assumption of linearity. Furthermore, the standardized residuals, when plotted against the model's predicted values, provide no evidence of the need for any transformation of the variables. This result helps justify the decision not to use the log transformation of the proportion poor as the dependent variable.

Constancy Over Time

Table 6-7 shows the regression coefficients for the predictor variables for the state model for each of the years from 1989 to 1996, excluding 1994. The coefficients for all four poverty-rate predictor variables are positive in all 7 years and generally similar across all years. All of the coefficients are significant at the 5 percent level except that the coefficient of the proportion of people under age 65 who were not included on an income tax return (column 3) is not significant in 1989.

Inclusion or Exclusion of Predictor Variables

The standardized residuals for the state regression model were grouped into four categories for each of the following characteristics: census region, population size in 1990, 1980 to 1990 population growth, percentage of black population in 1990, percentage of Hispanic population in 1990, percentage of group quarters residents in 1990, and percentage of poor school-age children in 1979 (from the 1980 census). The distributions of the standardized residuals for each category were then displayed using box plots. For none of these box plots is there an obvious pattern to the standardized residuals across categories, with one exception: in 1989, 1990, 1991, and 1993, the model underpredicts the proportion of poor school-age children in the West Region (i.e., the model estimates are lower than the CPS direct estimates for this group of states). The Census Bureau experimented with adding a West Region indicator predictor variable to the model. The coefficient of this variable has a negative sign for all 7 years; however, it is significant for only 1991, 1992, and 1993. For those 3 years, the

TABLE 6-7 Estimates of Regression Coefficients for the 1995 State Model, Estimated for 1989-1993, and 1995-1996

	Predictor Variables[a]			
Year	(1)	(2)	(3)	(4)
1989	0.52 (.09)	0.71 (.20)	0.23 (.13)	0.71 (.34)
1990	0.46 (.09)	0.65 (.20)	0.42 (.15)	1.07 (.36)
1991	0.46 (.10)	0.52 (.21)	0.59 (.14)	0.84 (.37)
1992	0.41 (.10)	0.71 (.21)	0.42 (.13)	1.38 (.37)
1993	0.28 (.12)	1.14 (.25)	0.51 (.14)	1.24 (.39)
1995	0.57 (.12)	0.79 (.25)	0.32 (.13)	1.54 (.36)
1996	0.37 (.12)	0.97 (.26)	0.59 (.14)	1.02 (.36)

NOTES: All predictor variables are in terms of rates. Standard errors of the estimated regression coefficients are in parentheses.

[a]Predictor variables: (1) ratio of child exemptions reported by families in poverty on tax returns to total child exemptions; (2) ratio of people receiving food stamps to total population; (3) ratio of people under age 65 who were not included an income tax return to total population under age 65; (4) residual from a regression of poverty rates for school-age children from the prior decennial census (1980 or 1990) on the other three predictor variables.

model with the West Region variable performs better for states in the West Region. A further examination of the residuals from the state model without the West Region predictor variable for individual Western states reveals that the model fairly consistently underpredicts the proportion of poor school-age children in some Western states but just as consistently overpredicts the proportion of poor school-age children in other Western states. Further investigation is needed to explain these patterns.

Normality, Homogeneous Variances, and Outliers

The distribution of the standardized residuals from the state regression model shows some small degree of skewness, especially in the 1992 equation. However, the skewness does not appear sufficiently marked to be a problem. Also, the residual plots and the box plots of the distributions of the standardized residuals against the categories of states show little evidence of any heterogenous variance. Finally, there is no evidence of outliers from examination of the residual plots or displays of the distributions of the standardized residuals from the state regression model.

Model Error Variance

One problem in the state model concerns the variance of the model error (u_i in equation (2) in Chapter 4). In the state model, the variances of the sampling errors (e_i in equation (2)) are estimated directly from the CPS data using a generalized variance function. The total model error variance is calculated using maximum likelihood estimation. The result of this calculation is an estimate of zero for the model error variance in the equation for every year except 1993. This result, which implies (absent sampling variability) that the model gives perfect predictions of state poverty rates for school-age children, is not credible. In the shrinkage estimate, it produces a zero weight for the direct estimates even when those estimates are quite precise, as is the case for several large states in the CPS sample. Even a small model error variance can substantially change the weight on the relatively high-precision direct estimates when they are combined in a shrinkage procedure with the model estimates.

To evaluate the effects of using zero model error variance in the estimation, the panel examined tables that compared the model estimates of the proportion of poor school-age children to the CPS direct estimates by state for 1989-1993 and 1995-1996; as an illustration, Table 6-8 shows this comparison for 1995. This examination demonstrated two important points. First, there are some appreciable differences between the model estimates and the direct estimates. For example, for Mississippi in 1995, the difference is over 7 percentage points. Therefore, if a non-zero estimate for model error variance is produced, it might have important consequences for the state estimates of poor school-age children. Second, while there are some appreciable differences, the model estimates were within two standard errors of the direct estimates for almost all states in each year. The range of model estimates that exceeded that limit in either a positive or negative direction was from one state in 1992 to six states in 1996. (Mississippi's difference in 1995 was not statistically significant at the 5 percent level.) For no single state did the model estimates exceed two standard errors of the direct estimates for more than 3 of the 7 years for which the state model was estimated. (And this analysis ignores the variance of the model estimates, which means that

TABLE 6-8 CPS Direct Estimate and Regression Model Estimate of Percentage of School-Age Children in Poverty by State, 1995

State	CPS Direct Estimate (1)	Lower Confidence Bound on Direct Estimate (2)	Upper Confidence Bound on Direct Estimate (3)	State Model Regression Estimate (4)	Regression Estimate Minus Direct Estimate (4) – (1) (5)
Alabama	22.2	16.5	27.9	23.4	1.2
Alaska	6.3	1.6	11.1	10.9	4.5
Arizona	23.0	16.8	29.2	21.1	–1.9
Arkansas	21.4	14.0	28.7	24.0	2.6
California	22.5	19.4	25.7	21.5	–1.0
Colorado	9.4	5.1	13.8	11.8	2.3
Connecticut	15.6	7.3	24.0	12.6	–3.0
Delaware	15.6	8.3	23.0	12.8	–2.8
District of Columbia	30.2	17.9	42.4	33.8	3.7
Florida	21.1	16.8	25.4	20.7	–0.4
Georgia	14.8	8.2	21.3	21.4	6.7
Hawaii	14.1	7.9	20.3	11.9	–2.2
Idaho	15.4	9.9	20.9	12.7	–2.7
Illinois	19.4	14.6	24.2	15.7	–3.7
Indiana	12.9	9.0	16.8	12.6	–0.4
Iowa	15.2	8.9	21.4	11.2	–3.9
Kansas	10.6	4.8	16.4	12.7	2.1
Kentucky	18.9	13.4	24.4	22.9	4.0
Louisiana	24.2	15.6	32.9	28.0	3.8
Maine	10.7	4.1	17.4	13.8	3.1
Maryland	12.8	5.0	20.5	11.5	–1.3
Massachusetts	16.5	11.5	21.5	13.3	–3.2
Michigan	14.2	10.0	18.3	17.2	3.0
Minnesota	9.5	5.5	13.4	10.0	0.6
Mississippi	34.9	25.6	44.3	27.4	–7.6
Missouri	9.4	3.5	15.2	17.0	7.7
Montana	17.4	9.4	25.3	18.4	1.0
Nebraska	11.4	7.1	15.7	10.0	–1.4
Nevada	9.8	4.0	15.6	11.8	2.0
New Hampshire	4.2	0.6	7.8	6.5	2.3
New Jersey	9.3	6.5	12.0	12.3	3.0
New Mexico	34.0	27.8	40.3	28.6	–5.5
New York	22.7	19.1	26.3	23.1	0.4
North Carolina	19.7	13.8	25.5	17.1	–2.6
North Dakota	10.3	5.3	15.2	14.1	3.8
Ohio	16.6	11.1	22.2	15.1	–1.5
Oklahoma	22.6	13.1	32.1	22.5	–0.1
Oregon	12.5	7.1	17.9	12.4	–0.1
Pennsylvania	16.1	12.5	19.7	15.3	–0.9
Rhode Island	16.4	10.7	22.2	15.1	–1.3
South Carolina	30.8	21.9	39.7	21.9	–8.9

TABLE 6-8 Continued

State	CPS Direct Estimate (1)	Lower Confidence Bound on Direct Estimate (2)	Upper Confidence Bound on Direct Estimate (3)	State Model Regression Estimate (4)	Regression Estimate Minus Direct Estimate (4) – (1) (5)
South Dakota	16.7	8.7	24.8	17.3	0.6
Tennessee	18.4	9.1	27.7	18.7	0.3
Texas	22.4	19.3	25.5	24.3	1.9
Utah	7.3	3.9	10.8	7.5	0.2
Vermont	11.3	3.2	19.4	11.6	0.3
Virginia	14.3	7.6	21.1	14.5	0.1
Washington	15.8	7.9	23.7	12.4	–3.4
West Virginia	23.0	13.2	32.9	25.7	2.7
Wisconsin	11.1	4.0	18.1	12.2	1.2
Wyoming	10.5	6.3	14.7	12.2	1.7

NOTE: Confidence bounds are plus or minus two standard errors on the direct estimate (95% confidence interval, obtained using direct estimates of the CPS standard errors).

SOURCE: Data from U.S. Census Bureau.

a yet smaller number of differences are statistically significant.) These results suggest that the state model is performing reasonably well: differences between model and direct estimates are neither unusually large nor strongly persistent. However, more work should be conducted to evaluate the current procedures for estimating the sampling error variance of the state model and the effects on the model estimates.

1990 Census Comparisons

Fay and Train (1997) compare 1989 estimates of the proportion of poor school-age children from the state model with 1990 census estimates. They find that the differences between the model and census estimates are much smaller than the differences between the 1989 CPS direct estimates and the 1990 census estimates and considerably smaller than the differences between the 1980 census estimates and the 1990 census estimates. These findings, which are presented graphically in Fay and Train (1997), support the use of a model-based approach to producing updated state estimates of poor school-age children instead of relying on estimates from the previous census or from the CPS alone. Similarly, a formal hypothesis test performed for the state model (Fay, 1996) supports the conclusion that the model-based estimates for 1993 are preferable to estimates from the 1990 census.[22] Comparable evaluations have not been performed for alternative state models or for categories of states.

State Raking Factors

The final stage in producing updated estimates of the number of poor school-age children for counties is to ratio adjust, or rake, the estimates from the county model for consistency with the estimates from the state model. The county model-1990 census comparisons found that the raking procedure was beneficial to the county estimates. The raking factors vary considerably across states. For 1995, the raking factors range from 0.71 to 1.14 (two-thirds fall between 0.88 and 1.06); for 1993, the raking factors range from 0.91 to 1.31 (two-thirds fall between 0.98 and 1.16).

The Census Bureau determined that the correlation between the raking factors for states in 1993 and 1995 is low, which implies that there is little systematic variation by state across these years. Also, some variation in the raking factors is expected given the form of the county model and the need to transform the predicted log values of poor school-age children to estimated numbers before the raking is performed. Other sources of this variability could include the use of 3-year averages of CPS estimates as the dependent variable in the county model versus single-year estimates in the state model, sampling variability, and, possibly, individual state effects that are not captured in the county model (see Chapter 9 and National Research Council, 2000:Ch.3). Preliminary work by the panel suggests that a large proportion of the variation in the state raking factors is due to sampling variability. Further investigation should be carried out to better understand the causes of this variation.

[22]The test assumes that the objective is to predict poverty rates that reflect the CPS measurement of poverty and not the decennial census measurement.

7

School District Estimates

For Title I fund allocations to be made in spring 1999 for the 1999-2000 school year, the Census Bureau was charged to produce updated estimates of the number of poor school-age children at the school district level. Three sets of school district estimates were required: (1) estimates of related school-age children (aged 5-17) who were in poor families in the preceding calendar year;[1] (2) estimates of all school-age children; and (3) estimates of the total population of the district. The first two sets of estimates were needed to implement the allocation formulas for basic and concentration grants; the third set of estimates was needed to determine which school districts have fewer than 20,000 people.[2]

This chapter considers estimates of poor school-age children for school districts. It reviews the difficulties that confront attempts to develop such estimates; describes the procedure that the Census Bureau used to develop district-level estimates of school-age children in July 1996 who were in poor families in 1995; and assesses the limited evaluations that are possible of these estimates. Finally, the chapter discusses the implications of the evaluations for the use of updated school district estimates for Title I allocations. Chapter 8 describes the procedure

[1]See Chapter 1 for the definition of related children.

[2]States, at their discretion, may aggregate the fund allocations for districts with less than 20,000 population and redistribute the funds by using another method that is approved by the Department of Education.

and evaluations for estimates of the number of all school-age children and of the total population in July 1996 for school districts.

ISSUES IN ESTIMATING POVERTY FOR SCHOOL DISTRICTS

Developing estimates of the number of poor school-age children (or other characteristics) for school districts presents difficult problems. These problems include the small population size of most districts and several other features of their boundaries and scope: school district boundaries in many instances cross county lines; they can and often do change over time; and some school districts cover specific grade levels, such as kindergarten-8 or 9-12. Because of these problems, there are no data sources now available for developing updated school district estimates of poor school-age children by using the type of model-based approach that was used for county estimates. These problems also compromise the quality of the estimates for school districts that can be made by aggregating data for blocks from the decennial census. We briefly review each of these issues in turn.

Size

Table 7-1 shows the distribution of total school districts, school districts coterminous with counties, and total counties by population size from the 1990 census. Of 15,226 districts, 49 percent had fewer than 5,000 people, and fully 82 percent had fewer than 20,000 people, while only 9 percent had 40,000 or more people; the median population size was about 5,250. By comparison, of 3,141 counties, 10 percent had fewer than 5,000 people, and 32 percent had 40,000 or more people; the median population size was about 23,000. Small districts, while numerous, accounted for small proportions of school-age children: districts with fewer than 5,000 people included only 6 percent of all school-age children, and districts with fewer than 20,000 people included only 27 percent of all school-age children; in contrast, districts with more than 40,000 people included 58 percent of all school-age children. Such uses as Title I fund allocations, however, require estimates for all school districts, no matter how small. Yet it is not possible to obtain direct estimates for school districts from national surveys, such as the March CPS. Many school districts will have no sampled households in national surveys, and the estimates for all but the largest districts with sampled households will be very unreliable (i.e., exhibit high sampling variability).[3] Even long-form census data, as discussed below, are unreliable for many school districts.

[3]The American Community Survey that is planned to start in 2003 will collect data from about 3 million housing units each year on an ongoing basis using an unclustered design. It will have sampled households in all school districts, but the sample size will not be large enough to produce sufficiently reliable estimates of poor school-age children for most districts even when the sample is aggregated over 5 years. See National Research Council (2000:Ch.4) for details.

TABLE 7-1 Percentage Distribution of School Districts, School Districts Coterminous with Counties, and Counties by Population Size, 1990 Census

	All School Districts		School Districts Coterminous with Counties		
Total Population	Districts (1)	School-Age Children (2)	Districts (3)	School-Age Children (4)	Counties (5)
Under 5,000	49.2	6.0	9.3	0.4	9.5
5,000-9,999	17.0	7.7	17.4	2.4	14.5
10,000-19,999	15.6	13.4	27.3	7.1	22.5
20,000-39,999	9.7	15.4	22.4	11.3	21.7
40,000 or more	8.5	57.6	23.7	78.8	31.7
Total (Number)	15,226	45.3 million	928	10.1 million	3,141

NOTE: School districts are defined as of 1989-1990.

SOURCE: Data from U.S. Census Bureau.

Boundaries

School district boundaries are, in general, determined by state regulations and practices. In seven states and the District of Columbia, school districts are coterminous with counties; these states included 370 districts in 1990 (2% of the total).[4] In another 17 states, school district boundaries coincide with other political units, such as townships. The boundaries of most, but not all, of the school districts in these states respect county lines. These states included 3,344 districts in 1990 (22% of the total), of which 190 crossed county lines. In the remaining 26 states, school district boundaries are unique to districts and often cross county lines. These states included 11,563 districts in 1990 (76% of the total), of which 3,931 crossed county lines. In all, 4,121 school districts (27% of the total) crossed county lines.

It is relatively easy to develop updated estimates of poor school-age children for districts that are coterminous with counties because county boundaries are generally stable over time, counties are relatively large areas, and data sources are available for counties (e.g., the data used to estimate the county model).

[4]In some other states, some school districts are coterminous with counties; see below. Puerto Rico is treated as a single county and (coterminous) school district for purposes of Title I allocations (see Appendix E).

Overall, in 1990, there were 928 districts that comprised an entire county or, in the case of a few districts (e.g., New York City), two or more entire counties. (The 928 districts include the districts in the seven states and the District of Columbia in which all school districts are counties together with selected districts in other states.) These districts accounted for 6 percent of all districts and 22 percent of all school-age children in 1990. Their median population size in 1990 was about 18,500 (Table 7-1, col. 3)–close to the median population size for all counties (Table 7-1, col. 5).

Most of the remaining districts, whether or not they cross county lines, present more or less serious problems for updating: they are small, with a median population size of less than 5,000; their boundaries can and often do change; and few data are available for estimating poverty. These districts accounted for 94 percent of districts and 78 percent of all school-age children in 1990.

Grade Levels

In 1990, 11,284 school districts (74% of the total) served all grades–pre-kindergarten, kindergarten, or 1st grade through 12th grade. The remaining 3,942 districts (26% of the total) served a subset of grades, such as elementary grades, high school grades, or middle school grades. Developing updated estimates of poor school-age children for districts that serve specific grades is difficult because a method must be devised to allocate the limited available data on school-age poverty to the age range that is appropriate to the grade range of the school district.

Data Sources

The Census Bureau's county model can readily provide updated estimates of the number of poor school-age children for the small subset of school districts that comprise entire counties. However, as noted above, a model similar to the county model cannot be developed for the remaining 94 percent of school districts, principally because of the lack of administrative data with which to form the predictor variables in a regression model. For example, states do not generally geocode the addresses of Food Stamp Program participants to school districts, so there are no counts of food stamp participants for school districts. Similarly, a substantial proportion of addresses on federal income tax returns cannot be geocoded to census blocks, so it is not possible to estimate the number of poor children reported by families on tax returns for school districts. Finally, data from school districts on students who are approved to receive free meals under the National School Lunch Program (requested from the states by the National Center for Education Statistics in its Common Core of Data program) are far from complete and are of uncertain quality and applicability (see below, "School Lunch Data"). In the future, it may be possible to develop appropriate data sources for a model-based approach to estimating poor school-age children for school districts (see Chapter 9; see also National Research Council, 2000: Ch.5), but such data are not now available.

ESTIMATION PROCEDURE

In the absence of data with which to develop a school district model similar to the county model, the Census Bureau used a simple within-county shares approach to estimate poor school-age children by school districts for 1995. The approach involved seven steps:

(1) A survey was conducted in which officials in every state were asked to provide school district boundaries for the 1995-1996 school year.

(2) Each 1990 census block was assigned to a school district, as defined for 1995-1996.[5]

(3) The 1990 census data were aggregated for the blocks (or fractions of blocks) in each school district or part of a school district that lay wholly within a county.

(4) The 1990 census data for each school district or school district part were tabulated to form a ratio estimate of the number of poor school-age children: the ratio estimate was obtained by applying the proportion of poor school-age children from the census long-form sample data to the short-form complete-count estimate of all school-age children. The ratio estimate was used because it reduced somewhat the high variability in the census estimates for school districts in comparison with estimates formed by simply inflating the long-form number of poor school-age children by the sampling weight.

(5) For the school districts or school district parts in a county, the share (proportion) for each school district or school district part of the 1990 census county total of poor school-age children was calculated from the ratio estimates. (For districts that are coterminous with a county, the share was 100%.)

(6) The 1990 census shares from step (5) were applied to the updated 1995 county estimates of poor school-age children produced by the county model (see Chapter 4) to obtain 1995 estimates of poor school-age children for school districts or school district parts.

(7) The 1995 school district estimates of poor school-age children were the

[5]When school district boundaries crossed census block boundaries, the poor school-age children in such a block were assigned to the appropriate school districts in proportion to the area of each district included in the block. When two or more school districts included a block because the districts covered selected grades (e.g., kindergarten-8 and 9-12), the poor children in the block in the relevant age ranges were assigned to the appropriate district on the basis of an analysis of the relationship of age to grade.

estimates from step (6) for school districts wholly within a county and the sum of the estimates of school district parts for school districts that crossed county lines.

As an example of the within-county shares procedure, take a county with 1,600 poor school-age children in 1989 (1990 census data) of whom 1,200 (75%) resided in school district A, 240 in school district B (15%), and 160 in school district C (10%). If the 1995 county model estimated that the county had only 1,200 poor school-age children, then the estimates of poor school-age children in 1995 for school districts A, B, and C are 900, 180, and 120, respectively. The estimation method assumes that all three school districts in the county experienced the same proportionate decrease in the number of poor school-age children–25 percent–as the county as a whole. If this assumption is incorrect (e.g., because the decrease in poverty in the county was concentrated in one of the districts, perhaps because of changes in the housing stock), then the estimates for the three school districts will be incorrect.

For the 1997-1998 school year, 18 states used a similar procedure for allocating their Title I county funds to school districts, in that they made within-county allocations on the basis of 1990 census school district shares of poor school-age children, either solely or in combination with estimates of the other categories of formula-eligible children (e.g., foster children). Another nine states used 1990 census data together with other data sources, such as school lunch data, to allocate Title I county funds to school districts (according to the U.S. Department of Education).

The Census Bureau's 1995 school district estimates of poor school-age children are not the only input to the Title I allocation formula. To make direct allocations to school districts for the 1999-2000 school year, the Department of Education also had to obtain several other data elements for school districts, most of which have not been previously available at the district level: counts of the other categories of formula-eligible children (children in foster homes, in local institutions for neglected and delinquent children, and in families with income above the poverty line who receive welfare assistance);[6] and the dollar amounts of Title I allocations that school districts received for the 1998-1999 school year (to use in the hold-harmless computations). The Census Bureau's estimates of poor school-age children also had to be adjusted to reflect school district boundary changes between 1995-1996 and 1998-1999; the department left it to the states to make appropriate adjustments.

[6]Poor school-age children estimated by the Census Bureau were 96.2 percent of the total number of formula-eligible children counted in the Title I allocations for the 1998-1999 school year. Foster children, children in local institutions for neglected and delinquent children, and children in families with income above the poverty line receiving welfare assistance were 2.6 percent, 1.1 percent, and 0.1 percent, respectively, of the total number of formula-eligible children.

EVALUATIONS

To evaluate the Census Bureau's 1995 estimates of poor school-age children for school districts, the panel and the Census Bureau first assessed the 1990 census estimates that are used to form school district shares of poor school-age children within counties. The 1990 census estimates are subject to high sampling variability, which is a problem for the Bureau's shares procedure. This high variability is also a problem for evaluations that use the 1990 census estimates as the standard of comparison.

Opportunities to evaluate the school district estimates are constrained by the limitations of available data. The panel and the Census Bureau used a 1980-1990 school district census file to evaluate a few variations of the Bureau's shares procedure for a subset of districts. The panel also evaluated the use of National School Lunch Program data as an alternative method for constructing updated school district estimates of poor school-age children in New York State.

Variability in Census Estimates

The two inputs to the Census Bureau's within-county shares procedure for school district estimates of the number of poor school-age children are the county model estimates for the target year, which have been extensively evaluated (see Chapter 6), and the 1990 census estimates for determining school district shares, which are discussed in this section. The income data that are used to determine poverty status in the census are collected on the long-form questionnaire, which was administered to an average of about one-sixth of households in 1990. The long-form sample size is orders of magnitude larger than the sample size of such household surveys as the CPS, but for small areas, the long-form estimates can exhibit high sampling variability.

Table 7-2 shows the mean and median coefficient of variation (in percent) for the estimated number of poor school-age children from the 1990 census long-form sample, obtained as a simple inflation estimate, for school districts distributed into groups categorized by number of school-age children, with each group containing approximately the same number of districts. The mean coefficient of variation is 32 percent for all school districts, varying from 64 percent for districts in the smallest size category (1-185 students) to 14 percent for districts in the largest size category (3,770 or more students).[7] This degree of variability is high. For example, if a typical school district has about 200 poor school-age children, the long-form sample might give estimates anywhere from about 70 to about 330 poor school-age children. (This range is from 200 minus twice the coefficient of variation of 32% for the typical district to 200 plus twice that

[7]The districts in the largest size category have about 20,000 or more total population.

TABLE 7-2 Average Coefficients of Variation (C.V.) for Two Estimates of Number of Poor School-Age Children for School Districts by Number of School-Age Children, 1990 Census

		Estimate from Long-Form Census Sample (in percent)		Estimate Ratio-Adjusted from Long Form and Short Form (in percent)	
Number of School-Age Children in District	Number of Districts	Mean C.V.	Median C.V.	Mean C.V.	Median C.V.
Total	14,328	32	23	30	22
1 to 185	1,858	64	54	57	47
186 to 462	2,446	39	30	36	28
463 to 946	2,480	32	24	30	22
947 to 1,811	2,505	28	22	26	21
1,812 to 3,769	2,519	23	19	22	18
3,770 or more	2,520	14	11	13	11

NOTES: Excludes school districts for which the estimated number of poor school-age children is zero. School districts are defined as of 1988-1990. The coefficient of variation is the standard error of the estimate divided by the estimate.

SOURCE: Data from U.S. Census Bureau.

coefficient of variation.) By comparison, a common design goal for estimates that are published from a survey is a coefficient of variation of 10 percent or less.

Table 7-2 also shows the mean and median coefficient of variation for school district estimates of poor school-age children that were constructed by ratio estimation. In this approach, the proportion of poor school-age children is computed from the long-form sample data and that proportion is then applied to the estimated total number of school-age children from the short-form or complete-count census data, which are not subject to sampling variability. This procedure somewhat reduces the variability of the estimates: the mean coefficient of variation of the ratio-adjusted estimates is 30 percent, compared with 32 percent for the long-form estimates, a reduction of 7 percent.

The Census Bureau used the ratio-adjusted 1990 census estimates of poor school-age children to construct the 1995 school district estimates but, given time constraints, did not conduct research on ways to further reduce the variability of the census estimates. One possible line of research is to use other short-form data (such as race and ethnicity, tenure, family type) as auxiliary information in the estimation of poor school-age children. Another line of research is to smooth the 1990 census school district estimates with the 1990 census county estimates, which would reduce the variability for smaller size districts (see Chapter 9).

Census Data Evaluations

The Census Bureau constructed a file of 1980 and 1990 census data for selected school districts, which was used to compare three sets of estimates of poor school-age children in 1989 with estimates from the 1990 census. In each instance, the 1980 census data that are used in the estimation are solely from the long form, while the 1990 census data are ratio adjusted. Three methods were used for the estimates:

(1) One method used county model estimates to construct school district estimates: method (1) applied the 1980 census shares of poor school-age children for school districts (or parts of school districts) within counties in 1979 to the Census Bureau's 1989 estimates of poor school-age children from its county model, with the county estimates controlled to the national estimate of poor school-age children in 1989 (from the 1990 census). This within-county shares procedure is analogous to that used by the Census Bureau to produce the 1995 school district estimates from 1990 census within-county shares applied to 1995 county model estimates, except that the 1980 census data are not ratio adjusted. (Also, the 1980 census estimates for 1979 are 10 years out of date for 1989 estimates, while the 1990 census estimates for 1989 are 6 years out of date for 1995 estimates.)

(2) A second method used 1990 census county estimates to construct school district estimates: method (2) applied the 1980 census shares of poor school-age children for school districts (or parts of school districts) within counties to the 1990 census county estimates of poor school-age children. This procedure eliminates the error in method (1) that is due to the county model.

(3) The third method was a national stable shares procedure: method (3) applied the 1980 census shares of poor school-age children for school districts within the nation as a whole to the national estimate of poor school-age children in 1989 from the 1990 census. This procedure assumes no change whatsoever in the relative shares of poor school-age children among school districts from the previous census, not even the change that occurs in methods (1) and (2) because of changes in the relative shares of poor school-age children among counties.

For several reasons, these comparisons provide only limited information with which to evaluate the Census Bureau's within-county shares model for school district estimates. First, the alternative models are not very different from the Census Bureau's model. Second, the 1990 census estimates that are the standard of comparison are subject to high sampling variability even after ratio estimation. Finally, the evaluation file, of necessity, contains only a subset of school districts.

Scope of Evaluation File

The 1980-1990 evaluation file was constructed from school district data sets that were prepared after each census. It was not possible to retabulate the individual block records from the 1980 census to match the 1990 census school district boundaries; instead, the goal was to identify a set of school districts in the data set for each year that could reasonably be assumed to have retained the same boundaries and grade ranges. The 1980 and 1990 census school district files were matched, using their identification numbers and other characteristics, and the following kinds of 1990 districts were dropped from the evaluation file:

- 928 districts or district parts for which the district or part was coterminous with a county and, hence, for which the county model would provide estimates;
- 4,108 districts that were not "unified," that is, that covered a limited grade range, such as kindergarten-8 or 9-12;
- 416 districts that were newly formed and had no counterpart in 1980;
- 12 districts in counties that changed boundaries between 1980 and 1990; and
- 609 districts that crossed county lines and for which one or more of the county pieces had no counterpart in 1980.

The resulting evaluation file contains 9,243 districts, which are 61 percent of the 15,226 school districts that were included in the 1990 census school district file and 56 percent of school-age children. The subset of school districts in the evaluation file closely resembles the entire set of 1990 school districts in terms of the distribution of total population and total number of school-age children in 1990. For example, the subset of districts in the evaluation file includes 47 percent with fewer than 5,000 people and 8 percent with more than 40,000 people; the corresponding figures for the entire set of 1990 school districts are 49 percent and 9 percent, respectively.

A key assumption for using the evaluation file is that the 9,243 districts in the file, which had the same identification numbers in both 1980 and 1990, are the

[8]Another assumption for using the evaluation file is that school districts for which the boundaries did not change from 1980 to 1990 represent the behavior of districts for which the boundaries did change. To the extent that changes in boundaries are associated with changes in population, the within-county shares approach may work less well for districts for which boundary changes occurred. However, these districts were less than 7 percent of the districts in 1990.

same districts and that their boundaries have not changed.[8] This assumption could be incorrect in some instances. For example, if a school district follows township boundaries and the township annexed land from another town between 1980 and 1990, it is likely that the school district identification number was the same in both 1980 and 1990 even though the boundaries changed.

To investigate this assumption, the Census Bureau looked at unified school districts, not coterminous with counties, that had the same identification numbers in 1990 and in the 1995-1996 school district boundary survey. For 6 percent of these districts, which accounted for 2 percent of school-age children, the total number of school-age children originally tabulated in the 1990 census differed by 5 percent or more from the number retabulated according to the 1995-1996 boundaries. For the remaining 94 percent of districts, the two tabulations were exactly the same or differed by less than 5 percent, indicating that the same identification number is a reasonably good indicator of stability in school district boundaries.

Summary of Evaluation Results: Absolute Differences

Table 7-3 provides summary statistics for the three sets of school district estimates of poor school-age children in 1989 in comparison with the 1990 census estimates. The statistics provided are the average absolute difference between the estimates from a model or method and the census, as a percentage of the average number of poor school-age children in the census, and the average proportional absolute difference between each set of estimates and the 1990 census estimates. For comparison purposes, the last row of the table provides the same statistics for county estimates of poor school-age children in 1989 from the Census Bureau's county model.

The first measure in Table 7-3 assesses the absolute difference between estimates from a method and the 1990 census in terms of numbers of poor children, while the second measure assesses the absolute difference in terms of proportional errors for school districts. From a national perspective, it can be argued that the absolute differences in terms of numbers are more important for effective Title I allocations because, with direct allocation, Title I funds are primarily distributed in proportion to the number of children in a school district. Therefore, the amount of funds that are misallocated depends primarily on the number of children rather than on the percentages by district (see Chapter 6). However, from the district perspective, the proportional error for a district's allocation is also important.

Ideally, a method will perform well on both types of measures, but, as discussed below, all three shares methods perform much worse on the average proportional absolute difference measure overall than on the average absolute difference measure. The reason for this consistent finding is that there are many small school districts that tend to have much larger-than-average proportional errors, which are reflected in the average proportional absolute difference mea-

TABLE 7-3 Comparison of Within-County Shares Estimates and 1990 Census School District Estimates of the Number of Poor School-Age Children in 1989

Model	Average Absolute Difference, Relative to Average Poor School-Age Children (in percent)[a]	Average Proportional Absolute Difference (in percent)[b]
1989 School District Estimates		
(1) Within-county shares method using 1980 census shares applied to 1989 county model estimates	22.2	60.0
(2) Within-county shares method using 1980 census shares applied to 1990 census county estimates	18.0	55.4
(3) National stable shares method using 1980 census shares applied to 1990 census national estimate	28.7	71.7
1989 County Estimates from Census Bureau's County Model	10.7	16.4

NOTES: School district estimates are based on 8,810 districts (9,243 districts in the 1980-1990 evaluation file minus 66 districts with estimated sample population of 30 or less in 1980 or 1990 and an additional 367 school districts with estimates of no children in poverty). The 1990 census estimates used in the comparisons are the ratio-adjusted estimates (see text). All three sets of school district estimates are controlled to the 1990 census national estimate of poor school-age children in 1989 before comparison with the 1990 census school district estimates.

[a]The formula, where there are n school districts or counties (i), and Y is the estimated number of poor school-age children from a model or the census, is

$$\Sigma[(|Y_{\text{model } i} - Y_{\text{census } i}|) / n] / [\Sigma(Y_{\text{census } i}) / n] .$$

[b]The formula is $\Sigma [(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}] / n$.

SOURCE: Data from U.S. Census Bureau.

sure. However, the much larger proportional errors for small districts do not represent many poor school-age children and so do not contribute as much to the absolute difference measure.

As seen in the last row of Table 7-3, the average absolute difference of the county model estimates from the 1990 census county estimates is 10.7 percent of the 1990 census county average number of poor school-age children; the average proportional absolute difference is 16.4 percent. The school district estimates show much larger differences. The average absolute difference for the Census Bureau's within-county shares method (1), which applies 1980 census school district shares of poor school-age children within counties to the county model estimates for 1989, is 22.2 percent of the 1990 census school district average number of poor school-age children (2.1 times the corresponding figure for the county model estimates); the average proportional difference is 60 percent (3.7 times the corresponding figure for the county model estimates).

Method (1) reduces the average absolute difference measure by 23 percent (22.2/28.7) and the average proportional absolute difference measure by 16 percent (60.0/71.7) compared with the national stable shares method (3), which assumes no change in school district shares of all poor school-age children in the nation between the 1980 and 1990 censuses. Method (2), which applies 1980 census school district shares within counties to the 1990 census county estimates of poor school-age children, performs somewhat better: it reduces the average absolute difference measure by 37 percent (18.0/28.7) and the average proportional absolute difference measure by 23 percent (55.4/71.7) when compared with the national stable shares method (3). However, method (2) is of theoretical interest only. In a noncensus year, such as 1995, model-based county estimates have to be used for adjusting school district shares from the census, and there will be errors in these estimates.

The Census Bureau also explored a fourth method in which a set of estimates was constructed by applying the 1980 census shares of poor school-age children for school districts within each *state* to the 1990 census state estimates of poor school-age children. This method produced average absolute and average proportional absolute differences between those of methods (2) and (3). It also is of theoretical interest only because it cannot be used in a noncensus year. However, it illustrates that using state estimates to control school district shares (which could be done with the Census Bureau's state model estimates) is better than using a single national control, but worse than using county controls.

There are several reasons for the large differences between the estimates of poor school-age children for districts produced by method (1) and the comparison ratio-adjusted estimates from the 1990 census: the sampling variability in the 1980 census estimates of school district shares, which is high for many districts; the inability of the within-county shares method to capture within-county changes in school district shares of poor school-age children from the 1980 census to the 1990 census; the errors in the county model itself (although these are not a large component); and the sampling variability that remains in the 1990 census comparison estimates even after ratio estimation. Because of the sizable sampling variability in the 1990 census estimates, the difference measures in Table 7-3 are overestimates of the differences from the true numbers of poor school-age children in 1989. It would be useful to remove this effect.

As an extension of this analysis, the Census Bureau has produced graphs of three quantities: a measure of the difference between the school district estimates from the census estimates, which is the root mean square difference; the estimated sampling variability of the census estimates; and the resulting calculated root mean square error of the school district estimates and the census estimates adjusted for the sampling variability in the latter (Bell et al., 2000). The graphs indicate that for school districts with small population sizes and small proportions of poor school-age children the sampling variability in the census estimates accounts for a sizable proportion of the root mean square difference. Extensions of this type of analysis to other categorizations of school districts would be useful.

Considering school districts by size, method (1) performs reasonably well on both the average absolute difference measure and the average proportional absolute difference measure for districts with 40,000 or more people in 1990 (data not shown). For these districts, the estimates are not markedly worse than the county estimates. Districts with 40,000 or more people are only 8 percent of the total number of school districts in the 1980-1990 evaluation file, but they contain 55 percent of the poor school-age children in the file.

Method (1) performs less well for school districts with 10,000 to 39,999 people in the 1990 census and performs very poorly for districts with fewer than 5,000 people in the 1990 census. Thus, while the average absolute difference measure for districts with 40,000 or more people in 1990 is 17 percent, it is 24 percent for districts with 20,000 to 39,999 people, 26 percent for districts with 10,000 to 19,999 people, 30 percent for districts with 5,000-9,999 people, and 43 percent for districts with 5,000 or fewer people. Districts with 5,000 or fewer people in 1990 contain only 8 percent of the poor school-age children in the 1980-1990 evaluation file, but they are 47 percent of total districts.

The much larger differences between the estimates from method (1) and the 1990 census estimates for smaller school districts relative to larger districts are due in part to the greater variability in the 1990 census estimates for smaller districts. As noted above, the panel believes there are ways to further reduce the variability in the 1990 census estimates beyond the reduction achieved by using simple ratio estimates instead of simple inflation estimates. A reduction in the variability of the 1990 census estimates would permit not only a more accurate assessment of the within-county shares approach, but also an improvement in the 1995 school district estimates that are formed by applying 1990 census within-county school district shares to the 1995 estimates from the county model.

Summary of Evaluation Results: Algebraic Differences

The evaluation also examined the algebraic differences by category of school district. The following categories were used: census geographic division, 1980 population, 1990 population, 1980-1990 population growth, percentage of poor school-age children in 1980, percentage of poor school-age children in 1990, change in the poverty rate for school-age children from 1980 to 1990, percentage of Hispanic population in 1980, percentage of black population in 1980, and percentage of group quarters residents in 1980. The results are summarized below for method (1); detailed results are provided in U.S. Census Bureau (1998b).

The category algebraic difference is the sum, for all school districts in a category, of the algebraic (signed) difference between the estimate of poor school-age children from a model or method and the 1990 census estimate for each district, divided by the sum of the census estimates for all districts in the category. This measure expresses model-census differences in terms of the numbers of poor children, similar to the overall absolute difference in the first column of Table 7-3. However, the category algebraic difference is expressed as an algebraic measure in which positive differences (overpredictions) within a category offset negative differences (underpredictions). The measure is intended to identify instances of potential bias in the predictions from a model or method. For example, the method may over(under)predict, on average, the number of poor school-age children in larger school districts relative to smaller districts.

The comparison of category algebraic differences for estimates from the Census Bureau's within-county shares method (1) with 1990 census estimates found no strong patterns of over(under)prediction for school districts categorized by percentage of black, percentage of Hispanic, or percentage of group quarters residents in 1980. However, method (1) did somewhat overpredict the number of poor school-age children in districts with no black or Hispanic residents or a very small proportion of group quarters residents in 1980 relative to other districts. Method (1) also somewhat overpredicted the number of poor school-age children in districts with fewer than 5,000 people in 1980 and 1990 relative to other districts. These findings may be related, in that districts with no black or Hispanic residents or very few group quarters residents are also districts that have very small populations.

For school districts categorized by population growth from 1980 to 1990, method (1) overpredicted the number of poor school-age children in districts that experienced a decline in population of more than 10 percent and underpredicted the number of poor school-age children in districts that experienced an increase in population of more than 10 percent. The same pattern was even greater for districts categorized by change in the poverty rate for school-age children from 1980 to 1990. These findings are not unexpected in that the within-county shares method, by definition, will not reflect large increases or decreases in population or poverty for school districts except to the extent that the district increase or decrease parallels that of the county in which it is located.

For school districts categorized by percentage of poor school-age children, method (1) underpredicted the number of poor school-age children in districts that had a lower school-age poverty rate in 1980 relative to districts with a higher rate. In contrast, method (1) overpredicted the number of poor school-age children in districts that had a lower school-age poverty rate in 1990 relative to districts with a higher rate. These findings are also not unexpected. They are evidence of the so-called "regression to the mean" phenomenon, in which, due to sampling variability, school districts that have low estimates of school-age poverty rates in one year will tend to have higher rates in another year (other things being equal) and vice versa.

Finally, for school districts categorized by census geographic division, method (1) overpredicted the number of poor school-age children in districts in the Pacific Division and, to a lesser extent, in the Mountain Division relative to districts in other divisions. This finding is consistent with a similar finding for the 1989 county model estimates, which, in turn, was attributed to the state model.

SCHOOL LUNCH DATA

As noted at the beginning of the chapter, there is a lack of administrative data with which to estimate school-age poverty for school districts. Food stamp data are not generally available for districts, and federal tax return data at present cannot be reliably coded to school district in many areas. Another possible source of information on poverty for school districts is data from the National School Lunch Program, which provides free and reduced-price meals to qualifying children.

The Census Bureau decided that it could not use school lunch data in developing updated estimates of poor school-age children for school districts for two major reasons. First, there is no complete and accurate set of school lunch data for all school districts. The National Center for Education Statistics (NCES) obtains school lunch counts as part of its Common Core of Data (CCD) system, in which state educational agencies report a large number of data items for public school systems.[9] The school lunch data are not published and have not been a priority of NCES. The center does not follow up with states when there is no information provided for a school district or to evaluate the accuracy of the reports. Hence, the quality of the data is not established, and they are far from complete.

Files of school lunch data for 1990-1995 that NCES provided to the panel contain large numbers of missing and zero values. In some cases, missing data may be due to the fact that a school district no longer exists (e.g., it may have been combined with another district); however, most instances of missing data appear to be due to nonreporting by school districts. Zero values may be valid in many instances, but NCES staff indicated that missing data are sometimes reported as zero, and analysis supported this assessment. Also, while states are asked to report counts of students approved to receive free lunches, it appears that many states report the combined total number of students approved for free or reduced-price lunches, which have different income eligibility limits.

Only 18 states have reports that are more than 90 percent complete (fewer

---

[9]NCES is the only federal agency that attempts to obtain school lunch data for school districts. The Department of Agriculture collects school lunch data but only aggregate counts at the state level. Each October it obtains state counts of the number of children approved for free lunch and reduced-price lunch in both public and participating private schools, and each month it obtains state counts of the number of meals served for purposes of reimbursing the states for meal costs (the subsidy varies by whether the meal was free, reduced price, or full price).

than 10% of school districts with missing or zero values) in all 6 years of the NCES files. At the other extreme, 10 states have reports that are less than 50 percent complete in all 6 years; most of these states do not report school lunch data at all. Clearly, if school lunch data are to be used to estimate the number of poor school-age children, it would be necessary to make school lunch reporting a priority in the CCD system for follow-up and evaluation.

The second reason for the Census Bureau not to use school lunch data in developing a consistent set of school district estimates nationwide is that counts of students approved for free lunches differ from poor school-age children in at least three respects, and the differences are probably not the same across jurisdictions:

- The eligibility standard to qualify for free lunches is family income that is less than 130 percent of the poverty threshold, which means that students approved for free lunches include near-poor as well as poor children. Children in families with incomes as high as 185 percent of poverty can receive reduced-price lunches.
- Participation in the school lunch program is voluntary and may be affected by such factors as perceived stigma (it is believed that high school students are less likely to participate than elementary school students for this reason) and the extent of outreach by school officials to encourage families to sign up for the program.
- Students approved for free lunches include children enrolled in participating schools in the district, whereas the Census Bureau is charged to produce estimates of poor school-age children who reside in the district. The two populations differ to the extent that poor resident children attend nonparticipating private schools or schools outside their district (nonresident poor children may also attend schools in the district).

If the relationship between students approved for free lunches and poor school-age children varies across jurisdictions, then it would not be possible to use school lunch data to estimate school-age poverty for school districts directly (e.g., by applying a constant factor to the school lunch counts to obtain estimated numbers of poor school-age children). If school district estimates are obtained by suballocating or distributing county-level estimates, then school lunch data could be used in modeling the suballocation if the relationship between students approved for free lunches and poor school-age children is constant across school districts within counties. However, variations in the relationship within counties would be a problem for such modeling.

There are two other reasons that such modeling could be problematic if school lunch data appeared suitable to use in models for some but not all states and counties. First, there would be practical difficulties for the Census Bureau to collect the data and develop and evaluate different estimation procedures for different sets of school districts, even when it might be possible to improve the accuracy of the estimates in some cases. Second, if the use of different estimation procedures produced estimates with different biases across school districts, there could be a problem of equity for concentration grants. The reason is that, under direct allocations, the concentration grant allocation to one area can affect the allocations to other areas. This was not the case under the two-stage allocation process, in which states that used school lunch data (or another data source) to allocate concentration grant funds to school districts were constrained by the county allocations determined by the Department of Education.

Yet the number of students approved for free lunches is an indicator of low income that relates specifically to the population of school-age children and that could be updated annually. Moreover, it is not subject to the sampling error that is such a serious problem for the Census Bureau's estimation procedure that suballocates county-model estimates on the basis of sample data from the census long form. Further, school lunch data carry considerable face validity with local officials.[10]

Thus, if school lunch data were available and determined to relate in a reasonably consistent manner to school-age poverty across jurisdictions, the Census Bureau could consider using such data to modify its current estimation process (see National Research Council, 2000:Ch.5). For example, it could follow the practice of the states that previously used school lunch data, solely or together or with census data, to distribute the Department of Education's Title I allocations for counties to school districts under the two-stage procedure. In effect, these states used a shares approach for school district estimates that is similar to the Census Bureau's method, except that the district shares within counties were computed on the basis of contemporaneous school lunch data instead of 1990 census estimates of poor school-age children.

The panel undertook a limited evaluation of a school lunch-based shares approach in one state–New York–for which it was able to obtain complete free and reduced-price school lunch data for almost all public schools for 1989-1990 and assign them to school districts and counties.[11] There are 623 New York State school districts in the 1980-1990 evaluation file, or 7 percent of the total number

[10]Interviews with state officials conducted in spring 1999 found widespread use of school lunch data as a proxy measure for poverty in allocating state funds and suballocating federal funds to school districts (Midwest Research Institute, 1999). School lunch data also appeared more credible to some state officials than the Census Bureau's estimates for allocating Title I funds to school districts.

[11]This evaluation was carried out at the State University of New York-Albany by Dr. James Wyckoff, a member of the panel, assisted by Frank Papa; see Appendix D, which includes overall and category comparisons.

of districts in the file. The New York State districts in the evaluation file are somewhat larger than average, with a median population size in 1990 of about 9,000 compared with a median population size of about 5,250 for all districts.

The analysis compared three sets of estimates of poor school-age children in 1989 for school districts in New York State with estimates from the 1990 census. The methods used to develop the three sets differ only in the estimation of within-county school district shares: the Census Bureau's method (2), in which 1980 census within-county school district shares of poor school-age children were applied to 1989 county estimates from the 1990 census; a method in which 1989-1990 within-county school district shares of students approved to receive free lunches were applied to 1989 county estimates from the 1990 census; and a method in which 1989-1990 within-county school district shares of students approved to receive free or reduced-price lunches were applied to 1989 county estimates from the 1990 census.

Table 7-4 provides summary statistics for the three sets of school district estimates of poor school-age children in 1989 for New York State compared with the 1990 census estimates for these districts. The table includes the average absolute difference between the estimates from a method and the census, expressed as a percent of the average number of poor school-age children in the census, and the average proportional absolute difference between each set of estimates and the 1990 census estimates. For comparison purposes, the last row of the table provides the same statistics for estimates of poor school-age children for all U.S. school districts in the evaluation file from method (2), which applies within-county school district shares to 1990 census county estimates.

The average absolute difference of the estimates for all school districts from the 1990 census estimates using method (2) is 18 percent; the average proportional absolute difference is 55 percent. The corresponding figures for estimates for New York State school districts only are 24 percent and 53 percent, respectively, for a method analogous to method (2); 22 percent and 49 percent, respectively, for a method based on free lunch counts; and 24 percent and 52 percent, respectively, for a method based on free and reduced-price lunch counts.

The absolute differences in all three methods of estimating poor school-age children in 1989 for New York State school districts are similar and large in magnitude. Even though the school lunch data pertain to the same year as the 1990 census comparison estimates, neither set of school lunch-based shares estimates is much more accurate than the 1980 census-based shares estimates. However, looking at both absolute differences and category algebraic differences, the use of free lunch counts as the basis for estimates is marginally more accurate than the other two methods that were evaluated. This finding suggests that it could be worthwhile to conduct a similar analysis for other states to determine if there is enough consistency across jurisdictions in the relationship of school lunch data to school-age poverty to warrant further consideration of the use of

TABLE 7-4 Comparison of Within-County Shares Estimates and 1990 Census School District Estimates of the Number of Poor School-Age Children in 1989, New York State

Model	Average Absolute Difference, Relative to Average Poor School-Age Children (in percent)[a]	Average Proportional Absolute Difference (in percent)[b]
New York State School District Estimates (N = 623)		
Within-county shares method (2) using 1980 census shares applied to 1990 census county estimates	23.9	53.4
Within-county shares method using 1989-1990 free lunch participants applied to 1990 census county estimates	22.3	48.7
Within-county shares method using 1989-1990 free and reduced-price lunch participants applied to 1990 census county estimates	24.2	52.1
U.S. School District Estimates (N = 8,810) from within-county shares method (2) using 1980 census shares applied to 1990 census county estimates	18.0	55.4

[a]The formula, where there are n school districts (i), and Y is the estimated number of poor school-age children from a model or the census, is

$$\Sigma[(|Y_{\text{model } i} - Y_{\text{census } i}|) / n] / [\Sigma(Y_{\text{census } i}) / n] .$$

[b]The formula is $\Sigma [(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}] / n$.

SOURCE: Wyckoff and Papa (in Appendix D); see also Table 7-3.

school lunch data for school district estimates.[12] If these data were to be used, a major effort would be needed to improve the reporting of the data to NCES for use by the Census Bureau for estimation purposes.

[12]A similar analysis was carried out for the state of Indiana at the University of Notre Dame by Dr. David Betson, a member of the panel (Betson, 1999). He assembled school lunch data for 1990-1991 for Indiana school districts. Difficulties in matching the school districts represented in the school lunch data set with the Census Bureau's set of school districts for Indiana prevented a full analysis. However, preliminary results were similar to the results from the New York State analysis–that is, estimates of within-county school district shares of poor school-age children in 1989 that were produced on the basis of 1980 census data and free lunch counts were roughly similar in accuracy when compared with 1990 census estimates of poor school-age children.

ASSESSMENT

It was difficult for the panel to draw firm conclusions from the evaluations of the Census Bureau's updated school district estimates of poor school-age children regarding their use for Title I allocations. On the positive side, the estimates are reasonably good for two groups of districts that contain many poor school-age children: districts that are coterminous with a county or more than one county, for which the county model provides estimates, and other districts with a total population of 40,000 or more, for which the Census Bureau's within-county shares method produces estimates that are only somewhat less reliable than the county model estimates.[13] These two groups together (adjusting for the overlap among them) comprise only a small fraction of districts, 13 percent of the total as of 1990, but they contain a large fraction of poor school-age children, 62 percent of the total. On the negative side, the school district estimates are highly variable for the remaining 87 percent of districts, which contain 38 percent of poor school-age children.

In terms of the mandate to the panel, the estimates might be judged to be "inappropriate or unreliable" for direct allocations of Title I funds to school districts. However, such a conclusion implies a definition of "inappropriate or unreliable" that does not take into account the allocation procedures that might otherwise be used. Given that some set of estimates will be used to make Title I allocations, the panel believed that "inappropriate or unreliable" should be defined in a relative sense. Applying a relative definition, one could argue that, in the context of currently available information, a direct allocation procedure that uses the Census Bureau's school district estimates is at least as good as and perhaps preferable to the alternative, which would be to return to the two-stage process in which the states distributed the county allocations from the Department of Education to school districts by using a variety of data sources.

As described in Chapter 2, the states used several types of data for suballocations of Title I funds when the two-stage procedure was in effect. For the 1997-1998 school year, 18 states relied on 1990 census data, either solely or together with estimates of the other categories of formula-eligible children, to distribute the county allocations to school districts. For these states, the Census Bureau's 1990 census shares-based estimates are likely to be somewhat more accurate than the corresponding estimates that the states were producing because the Bureau had access to 1990 census block data and so could more accurately retabulate the census data to reflect changes in school district boundaries; the states had access only to public use census files for 1989-1990 school district

[13] The 40,000 population size cutoff should be viewed as approximate. Examination of the evaluation results for a more detailed set of population size categories for school districts than discussed in the text indicated that the method (1) estimates for school districts approach the reliability of the county estimates somewhere in the range of about 30,000 to 50,000 population.

boundaries.[14] In addition, the ratio-adjustment procedure employed by the Census Bureau to estimate census shares somewhat reduces their variability. For the six states in this group that used 1990 census data to make direct allocations of basic grants to school districts without regard to the county allocation amounts, the use of the Bureau's 1990 census shares-based estimates has the advantage that they reflect the updated county estimates from the Bureau's county model.

Twenty-five states used data sources other than the census, or in combination with the census, to suballocate county Title I funds to school districts. (Three of these states made direct allocations of basic grants to districts.) It was not possible to evaluate the accuracy of such sources as school lunch data across states. The analysis that was conducted for New York (and the preliminary analysis for Indiana, see above) suggests that there are only marginal gains in accuracy from use of school lunch data. Moreover, it is not likely that the use of a shares approach based on school lunch data would produce results that are as consistent across states as the use of a shares approach based on census data: in some states, school lunch shares might be better than census shares; in other states, they might be worse. This inconsistency could be a problem for direct allocation of concentration grants.

Overall, the panel found four reasons to support use of the Census Bureau's school district estimates of poor school-age children for direct allocation of Title I allocation funds: the congressional mandate for direct allocations; the use of a uniform procedure to derive the Census Bureau's estimates; the somewhat greater accuracy of the Census Bureau's estimates of 1990 census shares compared with what the states could likely produce; and the absence of strong evidence that there are other, better data sources available for estimation. For the rest of the panel's assessment, it considered more carefully the features of the basic grant and concentration grant allocation formulas and how they might interact with the provision in the 1994 legislation that states may redistribute the aggregate allocations for districts with fewer than 20,000 people by some other method that the Department of Education approves.

Basic Grants

Under the two-stage allocation process, basic grants were allocated to school districts essentially as shares of the county total amounts. Whatever the data source used by a state to form the within-county shares (e.g., census data, school lunch data, combination of two or more data sources), the county totals remained as specified by the Department of Education. The exception is that the department allowed nine states in which school district boundaries bear little correspon-

[14]The Census Bureau provided the Department of Education with a file of 1990 census data for school districts defined according to 1995-1996 boundaries, to which the states can have access.

dence to county boundaries to redistribute the total basic grant allocation for the state without regard to the county allocations. For other states, the county totals, which, in turn, reflected (approximately) the Census Bureau's updated estimates from its county model, were maintained.[15]

Direct allocation of basic grants to school districts by using the Census Bureau's within-county shares estimates has the same property of essentially respecting the county totals because the Census Bureau's estimation procedure controls the school district estimates to county estimates derived separately from its county model. The correspondence between the county totals from the two-stage allocation process and those from the sums of direct allocations to the districts in each county will not be exact for several reasons. One, the hold-harmless provisions applied at the county level will give a somewhat different result from applying the hold-harmless provisions to districts and aggregating the resulting amounts to counties. Also, in contrast to counties, a proportion of school districts (about 10-12%) do not receive basic grants: although there was no eligibility threshold for counties to qualify for basic grants, school districts must have at least 10 formula-eligible children, and the number of eligible children must exceed 2 percent of the total number of school-age children in the district. Nonetheless, for basic grants, the county totals under direct allocations to school districts are likely to be fairly similar to what the allocations would be under the two-stage procedure.

However, when states choose the option in the legislation to redistribute the aggregate of the direct allocation amounts for school districts with fewer than 20,000 people by using some other data source (such as school lunch data), then the sum of the amounts for the districts in a county may not be similar to what the county amount would be under the two-stage process. The panel was concerned about this possible outcome: the county allocations that were made under the two-stage process reflected (approximately) the Census Bureau's county estimates from its county model, and these estimates are the only small-area estimates of poor school-age children that have been thoroughly evaluated and determined to be reasonably reliable.[16] Direct allocations that use the Census Bureau's within-county shares estimates for school districts also reflect (approximately) the Bureau's county estimates, but state plans to redistribute the direct allocation amounts for school districts with fewer than 20,000 people by using some other data source may not have this desirable property.

---

[15]The county allocations under the two-stage allocation process corresponded only approximately to the county model estimates because of other factors in the allocation formula, such as hold-harmless provisions.

[16]For example, the county estimates of poor school-age children developed from the county model are much more reliable than county estimates developed by such methods as applying within-state county shares of poor school-age children in the previous census to updated estimates from the Census Bureau's state model (see Chapter 6).

Analysis with 1989 school lunch data for New York State districts with fewer than 20,000 people (476 districts, see Appendix D, Table D-9) did not find evidence of this problem. The average absolute and average proportional absolute differences from 1990 census school district estimates of poor school-age children were about the same for estimates that were developed by using free lunch counts with and without county controls. However, this analysis pertains to only one alternate data source in only one state. In the absence of a complete analysis of alternate data sources, the panel believed it to be desirable, to the extent possible, that the basic grant allocations reflect the county model estimates in all states, including those that choose the option of redistributing the aggregate of the direct allocations for school districts under 20,000 population by using another data source. The Department of Education can achieve this outcome by approving state reallocation plans that, in general, propose to aggregate the direct allocation amounts for districts under 20,000 population within counties and redistribute the county totals among the districts under 20,000 population in each county.

Concentration Grants

Concentration grants, in contrast to basic grants, were never allocated as shares of the county totals under the two-stage procedure because only a fraction (less than half) of jurisdictions was eligible.[17] Under the two-stage process, concentration grants were allocated to those counties that had more than 6,500 or more than 15 percent of formula-eligible school-age children. In turn, states allocated county concentration grants to those districts in eligible counties that exceeded the threshold number or percentage of formula-eligible children: most districts that qualified for concentration grants presumably did so on the basis of exceeding the percentage threshold; few presumably did so on the basis of having more than 6,500 formula-eligible children.

Tabulations of 1990 census data in the evaluation file identified 30 percent of school districts, containing 60 percent of poor school-age children, as eligible for concentration grants under the two-stage allocation process.[18] Eligible districts under the two-stage process were 65 percent of the total districts in eligible counties. (In states that used another data source, such as free lunch counts, to distribute county concentration amounts to districts, a higher percentage of school districts in eligible counties were likely classified as eligible for concentration grants; see below.)

[17]In contrast, all counties and almost 90 percent of school districts were eligible for basic grants.

[18]The tabulations were limited to districts in the 1980-1990 evaluation file for which the boundaries did not cross county lines, totaling 6,434 districts, or 70 percent of the districts in the evaluation file.

The census tabulations showed that an additional 9 percent of school districts, containing 14 percent of poor school-age children, would be eligible for concentration grants except that they are located in a county that is not eligible. Under the two-stage procedure, states could reserve up to 2 percent of their concentration grant funds to allocate to eligible districts that were not in eligible counties, but these amounts were probably not adequate for the children in those districts.

The panel noted that the use of fixed thresholds for concentration grants places great demands on the quality of the estimates of those thresholds. An error of only one poor school-age child can make the difference between receiving a grant and not receiving a grant. For school districts that receive concentration grants to which they would not be entitled if true estimates of poor school-age children were available, these errors will be perpetuated through the hold-harmless provisions, particularly if the hold-harmless rate is retained at 100 percent. (There are also fixed thresholds for school districts to receive basic grants, although they are low, as noted above.)[19]

Evaluation

One of the reasons for the legislation mandating direct allocations to school districts was to target concentration grants to all eligible school districts, including those in ineligible counties. To assess the appropriateness and reliability of the Census Bureau's updated school district estimates of poor school-age children for direct allocation of concentration grants, the panel first examined the rate of agreement between the Census Bureau's within-county shares method (1) and the 1990 census in classifying school districts into one of two poverty rate categories for school-age children in 1989 that correspond to the concentration grant threshold: 0 to 15 percent and 15 percent or higher; see Table 7-5. The tabulations were prepared from the 1980-1990 evaluation file for districts that did not cross county lines.

The method (1) school district estimates and the 1990 census ratio-adjusted estimates for 1989 assigned the same poverty rate category (0 to 15% or 15% or higher) to 76 percent of school districts and 87 percent of poor school-age children. By comparison, the county model estimates and the 1990 census county estimates for 1989 assigned the same poverty rate category to 88 percent of counties and 92 percent of poor school-age children. The rate of agreement between the method (1) school district estimates and the 1990 census ratio-adjusted estimates was least for school districts with fewer than 5,000 people: 64

[19]For a discussion of issues in the relationship of funding formulas and data sources, see National Research Council (2000:App.).

TABLE 7-5 Agreement Between Within-County Shares Method (1) Estimates and 1990 Census School District Estimates for Proportions of School-Age Children in Poverty in 1989

Method of Estimate	Percentage of School Districts	Percentage of Poor School-Age Children
Method (1) and Census Estimate, All Districts		
Both under 15%	50.0	25.6
Both 15% or more	25.7	60.9
(Total in agreement)	(75.7)	(86.5)
Census under 15%, method (1) 15% or more	8.8	2.5
Census 15% or more, method (1) under 15%	15.6	11.0
Method (1) and Census Estimate, Districts Under 5,000 Population		
Both under 15%	37.6	20.2
Both 15% or more	26.6	44.9
(Total in agreement)	(64.2)	(65.1)
Census under 15%, method (1) 15% or more	14.1	6.4
Census 15% or more, method (1) Under 15%	21.6	28.5
Method (1) and Census Estimate, Districts of 40,000 or More Population		
Both under 15%	59.8	22.0
Both 15% or more	31.8	70.0
(Total in agreement)	(91.6)	(92.0)
Census under 15%, method (1) 15% or more	2.4	1.3
Census 15% or more, method (1) under 15%	6.0	6.8
County Model and Census Estimate, All Counties		
Both under 15%	30.5	40.9
Both 15% or more	57.1	50.7
(Total in agreement)	(87.6)	(91.6)

NOTES: School district estimates are based on 9,243 districts in the 1980-1990 evaluation file. The 1990 census estimates for school districts are the ratio-adjusted estimates (see text). The method (1) school district estimates are produced by applying 1980 census within-county school district shares of poor school-age children to the county model estimates for 1989 and controlling to the 1990 census national estimate of poor school-age children in 1989.

SOURCE: Data from U.S. Census Bureau; see Chapter 6, Table 6-5 (model b) for county model comparisons.

percent agreement for districts and 65 percent agreement for poor school-age children.[20] The rate of agreement was highest for school districts with 40,000 or more people: 92 percent for both districts and poor school-age children, slightly

[20]At least part of the explanation is that the census comparison estimates are subject to particularly high sampling variability for the smallest districts (see Table 7-2).

better than the rate of agreement for counties. For districts and poor school-age children for which the method (1) and 1990 census estimates were not in agreement, method (1) classified a higher percentage in the under 15 percent school-age poverty rate category.

To focus on the issue of concentration grant eligibility for school districts with direct allocations versus the two-stage process, the panel examined the correspondence between the method (1) estimates and the 1990 census estimates for cross-classifications of 1989 school district and county school-age poverty rate categories; see Tables 7-6 and 7-7. Method (1) estimated that 32 percent of districts, containing 59 percent of poor school-age children would be eligible for a concentration grant under the two-stage process (cell f, Tables 7-6 and 7-7), and that another 10 percent of districts, containing 12 percent of poor school-age children would be eligible for a concentration grant under direct allocations (cell o). These aggregate percentages are similar to those for the 1990 census, noted above (see cells h and q in Tables 7-6 and 7-7), but method (1) and the 1990 census classified a number of districts differently.

Of the districts and poor school-age children that the 1990 census estimated would be eligible for concentration grants under the two-stage process, method (1) agreed for 86 percent of districts and 96 percent of poor school-age children (cell e divided by cell h). The other 14 percent of districts and 4 percent of poor school-age children would be ineligible for concentration grants under the two-stage process according to method (1). There are also districts and poor school-age children that would be eligible under the two-stage process according to method (1) but ineligible according to the 1990 census: they comprise 18 percent of the districts and 3 percent of the poor school-age children that are eligible according to method (1) (cell d divided by cell f).

Of the additional districts and poor school-age children that the 1990 census estimated would be eligible for concentration grants under direct allocations (i.e., those in counties with school-age poverty rates under 15%), method (1) agreed for 53 percent of districts and 76 percent of poor school-age children (cell n divided by cell q). The other 47 percent of the additional districts and 24 percent of the additional poor school-age children would be ineligible according to method (1). There are also additional districts and poor school-age children that would be eligible according to method (1) but ineligible according to the 1990 census: they comprise 49 percent of the additional districts and 10 percent of the additional poor school-age children that are eligible according to method (1) (cell m divided by cell o).

Overall, the classification differences between the 1990 census estimates and the method (1) estimates are relatively large for the additional districts that would be eligible under direct allocations (i.e., districts with 15% or more poor school-age children in counties with less than 15% poor school-age children). However, the classification differences are relatively small for the additional poor school-age children that would be eligible under direct allocations. In particular, the

TABLE 7-6 Comparison of Within-County Shares Method (1) and 1990 Census School District Estimates for Proportions of School-Age Children in Poverty in 1989, by 1990 Census County School-Age Poverty Rate: Distribution by Percentage of School Districts

	Census School District Rate		
	Under 15%	15% or More	Total
CENSUS COUNTY SCHOOL-AGE POVERTY RATE 15% OR MORE			
Method (1) School District Rate			
Under 15%	10.8 (a)	4.3 (b)	15.1 (c)
15% or more	5.7 (d)	25.9 (e)	31.6 (f)
Subtotal	16.6 (g)	30.2 (h)	46.8 (i)
CENSUS COUNTY SCHOOL-AGE POVERTY RATE UNDER 15%			
Method (1) School District Rate			
Under 15%	38.9 (j)	4.4 (k)	43.3 (l)
15% or more	4.9 (m)	5.0 (n)	9.9 (o)
Subtotal	43.8 (p)	9.4 (q)	53.2 (r)
Total	60.4	39.6	100.0

NOTES: The two poverty rate categories used are those specified for concentration grants, 0-15 percent and 15 percent or more.

Cell entries are percentages of the 6,434 school districts in the 1980-1990 evaluation file for which the boundaries did not cross county lines. The 1990 census county and school district estimates are the ratio-adjusted estimates (see text). The method (1) school district estimates are produced by applying 1980 census within-county school district shares of poor school-age children to the county model estimates for 1989 and controlling to the 1990 census national estimate of poor school-age children in 1989. See text for discussion.

SOURCE: Data from U.S. Census Bureau.

TABLE 7-7 Comparison of Within-County Shares Method (1) and 1990 Census School District Estimates for Proportions of School-Age Children in Poverty in 1989, by 1990 Census County School-Age Poverty Rate: Distribution by Percentage of Poor School-Age Children

	Census School District Rate		
	Under 15%	15% or More	Total
CENSUS COUNTY SCHOOL-AGE POVERTY RATE 15% OR MORE			
Method (1) School District Rate			
Under 15%	6.1 (a)	2.5 (b)	8.6 (c)
15% or more	1.5 (d)	57.5 (e)	59.0 (f)
Subtotal	7.5 (g)	60.0 (h)	67.5 (i)
CENSUS COUNTY SCHOOL-AGE POVERTY RATE UNDER 15%			
Method (1) School District Rate			
Under 15%	17.3 (j)	3.3 (k)	20.6 (l)
15% or more	1.2 (m)	10.7 (n)	11.9 (o)
Subtotal	18.5 (p)	14.0 (q)	32.5 (r)
Total	26.0	74.0	100.0

NOTES: The two poverty rate categories used are those specified for concentration grants, 0-15 percent and 15 percent or more.

Cell entries are percentages of poor school-age children in 1989 in the 6,434 school districts in the 1980-1990 evaluation file for which the boundaries did not cross county lines. The 1990 census county and school district estimates are the ratio-adjusted estimates (see text). The method (1) school district estimates are produced by applying 1980 census within-county school district shares of poor school-age children to the county model estimates for 1989 and controlling to the 1990 census national estimate of poor school-age children in 1989. See text for discussion.

SOURCE: Data from U.S. Census Bureau.

percentage of poor school-age children in the additional districts that would be eligible for concentration grants according to the within-county shares method (1) estimates but would not be eligible according to the 1990 census estimates is relatively small (10%).

It should be kept in mind that these evaluations are limited in at least three ways. First, they apply only to a subset of school districts in the evaluation file, which are, themselves, a subset of total districts. Second, like all of the evaluations of the Census Bureau's school district estimates, they are based on a single comparison point. Third, the 1990 census estimates that are the standard of comparison are subject to high sampling variability for smaller school districts even with ratio adjustment.

Understanding the limits of the evaluations and the alternatives available, the panel concluded, on balance, that the use of the Census Bureau's school district estimates for direct allocations of concentration grants would be an improvement over the two-stage process. As intended by the 1994 legislation, many of the eligible districts that could not receive concentration grants with a two-stage allocation would receive such grants with direct allocations.

Reallocation of Concentration Grants

The option for states to redistribute concentration grant direct allocations for school districts with fewer than 20,000 people raised several issues for the panel to consider. States could propose to use another method, not only to redistribute the allocations among the districts that the Department of Education determined to be eligible for concentration grants on the basis of the Census Bureau's estimates, but also to redetermine eligibility. Under the previous two-stage procedure, the states that distributed county concentration grant allocations to districts on the basis of some other data source than the census used the alternate data source for both eligibility and amounts.

The use of free lunch or free and reduced-price lunch data in place of estimates of poor school-age children to redetermine eligibility as well as to redistribute allocation amounts likely has the effect that more districts receive concentration grants than would be the case with the use of the Census Bureau's school-age poverty estimates. The reason is that the income eligibility thresholds for free or reduced-price school lunches are higher than the poverty threshold. Consequently, more children fall below 130 percent of poverty (the threshold for free lunches) or below 185 percent of poverty (the threshold for reduced-price lunches) than fall below 100 percent of poverty.[21] (About 20% of school-age children nationally are in families with incomes below 100% of the poverty threshold, while about 26% are in families with incomes below 130% of the

[21]However, not all eligible children apply for reduced-price lunches.

poverty threshold and about 38% are in families with incomes below 185% of the poverty threshold.)[22] For this same reason, it is likely that proportionately more districts received concentration grants under the two-stage process in states that used school lunch data to determine eligibility than in states that used 1990 census data. In either case, the effect is to spread concentration grant dollars more thinly.

Analysis with 1989 school lunch data for New York State school districts with fewer than 20,000 people (476 districts; see Appendix D, Tables D-5 through D-8) provides evidence of the effect of using estimates that reflect higher poverty thresholds. Under the two-stage process, 136 such districts in New York State would be eligible for concentration grants by using free and reduced-price lunch data and 112 would be eligible by using free lunch data, whereas only 76 districts would be eligible according to the method (2) estimates (or the 1990 census). Under direct allocations, the effect is much more pronounced: 294 districts with fewer than 20,000 people would be eligible for concentration grants by using free and reduced-price lunch data, and 214 districts would be eligible by using free lunch data, whereas only 109 districts would be eligible according to the method (2) estimates (115 districts according to the 1990 census).

As noted above, the panel concluded that any redistribution of basic grant direct allocations for districts with fewer than 20,000 people should be performed for such districts within each county to the extent possible, thereby reflecting (approximately) the county estimates of poor school-age children. For concentration grants, the panel reached the same conclusion, although it should be noted that there may be a problem with this approach when different data are used for reallocation. For example, if a county has two school districts and only one district is eligible for a concentration grant according to the Census Bureau's estimates of poor school-age children, but both districts are eligible by using school lunch data, then the first district will lose some of its dollars to the second district. Presumably, similar situations occurred under the two-stage allocation process, in which school district concentration grants were allotted from county totals.[23] However, such situations may be somewhat more likely to occur under direct allocations, which provide concentration grants to eligible districts in counties that do not meet the concentration grant threshold.

One approach that could ameliorate this effect is to adjust school lunch data for school districts in a county to equal the Census Bureau's estimate of total poor school-age children for the county. The use of adjusted school lunch data to determine school-age poverty rates would be less likely to result in a much larger number of school districts qualifying for concentration grants than the use of the

[22]Data from panel tabulations of the March CPS for income years 1994-1996.

[23]The New York State analysis, in which more districts were eligible for concentration grants under the two-stage process by using school lunch data than by using the method (2) estimates, suggests that such situations occurred in the past.

Census Bureau's estimates of school-age poverty rates. Analysis conducted for New York State confirmed this outcome (see Appendix D, Tables D-7, D-8): 127 school districts with fewer than 20,000 people would be eligible for concentration grants under direct allocations by using adjusted free and reduced-price lunch data versus 294 districts that would be eligible by using unadjusted data. The corresponding figures are 124 districts and 214 districts by using adjusted and unadjusted free lunch data. By comparison, 109 districts would be eligible by using the method (2) estimates.

Study of the Allocation Process

Overall, by applying a relative standard for evaluation, the panel found reasons to support the use of the Census Bureau's updated estimates of poor school-age children for direct allocation to school districts. Also, the panel concluded that, in general, it is desirable for both basic grant and concentration grant allocations to reflect the county model estimates in all states, including those that choose the option of redistributing the direct allocations for school districts under 20,000 population by using another data source. However, the panel recognized that there are uncertainties about the operation of the formulas: for example, the extent to which the sum of direct school district allocations for counties approximates the allocations that would result for counties under the two-stage process and the extent to which there may be significant reallocations of concentration grant dollars from poorer to less poor districts with county controls. For this reason, the panel believed it to be critically important for the Department of Education to undertake a thorough study of the direct allocation process, both the methods used by the states and the results. Simulations of the allocations that would likely have been made under the two-stage process would be very helpful to inform the study.

8

Population Estimates

The SAIPE Program uses population estimates from the Census Bureau's postcensal population estimates program to form predictor variables in the state and county models of poor school-age children–the state population under age 65 in the state model and the county population under age 18 in the county model. In addition, state population estimates of noninstitutionalized children aged 5-17 are used to convert estimates of the proportion of poor school-age children from the state model to estimates of the number poor (see Chapter 4).

For the two-stage allocation procedure that the Department of Education used to allocate Title I funds prior to school year 1999-2000, the Census Bureau provided not only estimates of the number of poor school-age children in each county from the SAIPE Program, but also county estimates for the 5-17 age group to use as denominators for calculating the proportion of poor school-age children. (Estimates were also provided for Puerto Rico.) Both numbers and proportions are needed to determine eligibility and allocation amounts for basic and concentration grants (see Chapter 2).[1] The population estimates of school-age children that accompanied the 1993 county model estimates pertain to July

[1]The Census Bureau also makes available on its web site estimated proportions of poor school-age children in which the denominators are estimates of related children aged 5-17 in each county. These estimates are developed by multiplying the estimates from the Census Bureau's population estimates program for the noninstitutionalized population aged 5-17 by the ratio of related children aged 5-17 to noninstitutionalized children aged 5-17 for each county in the 1990 census.

1994; those that accompanied the 1995 county model estimates pertain to July 1996.

To enable the Department of Education to make direct allocations to school districts, the Census Bureau was charged to produce estimates at the school district level not only of poor school-age children in 1995, but also of the total population and total number of school-age children as of July 1996. Estimates of total school-age children are needed to compute poverty rates for school districts, which are a factor in the Title I allocation formulas. Estimates of total population are needed so that a state knows which districts have fewer than 20,000 people if it wants to take advantage of the provision in the legislation that permits states to aggregate the Title I allocations for these districts and to redistribute the funds on some other approved basis.

The Census Bureau currently develops county age estimates within the framework of total population estimates for counties and population estimates by age for states. School district estimates of total population and school-age children are developed by using a shares procedure, similar to that used for school district estimates of poor school-age children. In this procedure, 1990 census data for school districts are applied to updated county population estimates.

METHODS FOR POPULATION ESTIMATES

This section describes the methods that the Census Bureau used to develop the following population estimates: county estimates of total population for 1994 and 1996; county estimates of the population by age for 1994 and 1996; and school district estimates of total population and school-age children for 1996. The descriptions of methods for county estimates of total population and population by age briefly summarize the methods used for the corresponding estimates for states (for more detail, see Long, 1993; Sink, 1996; U.S. Census Bureau, 1995b).[2]

County Estimates of Total Population

In a process that begins anew with each decennial census, county estimates of total population are developed by updating the population estimates for the preceding year with data on births, deaths, net immigration from abroad, and net internal migration. The method is a component method, in which the numbers of births and deaths are based on reported vital statistics for each county; reports of

[2]Estimates for Puerto Rico are developed separately. The basic methodology for 1994 and 1996 estimates used registered births by sex, registered deaths by age and sex, and estimates of annual intercensal net migration by age and sex from an analysis using the natural rate of increase for the 1980-1990 period and the reported 1990 census population by age and sex (Reed, 1996).

the Immigration and Naturalization Service are used to estimate net immigration from abroad; and administrative records are used to estimate net migration among counties. Net migration of people under 65 years of age is estimated for each county from a year-to-year match of IRS federal income tax returns; for people aged 65 and over, net migration is estimated for each county from the change in Medicare enrollment (U.S. Census Bureau, 1995). Estimates are developed separately for household and group quarters populations.

The county population totals are summed for each state to provide estimates of the total population of each state. All county and state population totals are then adjusted to sum to independently derived estimates of the total U.S. population.[3] The county estimates are also reviewed locally under the Census Bureau's Federal-State Cooperative Population Estimates (FSCPE) Program.

Operationally, the county total population estimates are the sum of the estimates for four groups:

- Household population under age 65 (HHP < 65);
- Household population age 65 and over (HHP65+);
- Group quarters population under age 65 (GQ < 65); and
- Group quarters population age 65 and over (GQ65+).

Household Population Under Age 65 The estimates for the household population under age 65 use a component method for year t to measure the change in each component of population change during the 12-month period preceding the estimate date, as follows:

$$\text{HHP} < 65_t = \text{HHP} < 65_{t-1} + \text{NI} + \text{NMIG} + \text{NETMOVE} - \text{AGE}. \qquad (1)$$

NI is natural increase (births and deaths for people under age 65), which is estimated from a combination of vital statistics data from the National Center for Health Statistics (NCHS) and from state agencies that participate in the FSCPE Program. Each of these sources has some problems. The FSCPE does not

[3]The national-level population estimates are not adjusted for net undercount in the census, but the methodology includes an "inflation-deflation" procedure so that the undercount patterns for age groups are consistent between the estimates and the census. In this procedure the census counts for age groups are first adjusted (inflated) for net undercount as estimated from demographic analysis. Then the adjusted counts are carried forward by subtracting deaths, adding net immigration, and making the group 1 year older for each year of the estimates (births are added for the age group under 1 year). As a last step, the estimates are deflated, using the net undercount rates that apply to the updated age group. As an example, census counts for men age 20 are inflated using the relatively high net undercount rate for that age. After the updating is carried out over, say, 10 years, the resulting estimates are deflated by using the net undercount rate for men age 30, which is a smaller rate than the rate for men age 20.

include all states, and the NCHS data exhibit some peculiarities (e.g., birth records are not always properly assigned to place of residence in such areas as Washington, D.C., in which births often occur in hospitals that are not in the county of residence, and in areas with military bases).

NMIG, net internal migration, is estimated from data on IRS tax returns matched year to year on the basis of the social security number of the filer. A migration rate is developed from the net flow of exemptions (the tax filer and his or her dependents) on the matched tax returns. The rate is calculated as the difference in the number of exemptions entering the county minus the number leaving the county, as a proportion of the number of exemptions at the start of the period. This rate is then applied to the migration base [$HHP < 65_{t-1} + 0.5(NI + NETMOVE) - AGE$]. Coverage of the IRS data (i.e., the proportion of exemptions to estimated population) varies across counties, as do matching rates.

NETMOVE is nondomestic net movements, mainly international immigration and emigration. It is estimated with a variety of data, and the totals generally are small. *Legal immigrants and refugees* (about 800,000 per year nationwide) are assigned to a county of residence on the basis of Immigration and Naturalization Service data about their intended place of residence, although they may not reside at the indicated place. *Undocumented immigrants* (estimated at 225,000 annually) are assigned to a county on the basis of the 1990 census distribution of the foreign born population. Estimates are also made of *emigrants* (about 195,000 per year). *Net inmigrants from Puerto Rico* (only about 7,000 annually because there is almost an equal number of outmigrants each year) were previously estimated from passenger traffic data from the San Juan airport. However, this method became increasingly untenable, and the current procedure uses estimates of migration of Puerto Ricans to the rest of the world, which include an assumption of the U.S. share. The U.S. share is allocated to counties on the basis of 1990 census data on place of residence. Estimates of the net movement in and out of the country of *military and federal civilian and military dependents* are based on data from the Department of Defense (DoD) and the Office of Personnel Management. County station strength data from DoD, which are used to allocate military personnel to counties, are modified in some locations (e.g., the Washington, D.C., area).

Lastly, AGE is an estimate of the number of persons in the county who aged from 64 to 65 during the year.

Except for internal migration, all components are controlled to national totals.

Household Population Age 65 and Over The estimates for the household population age 65 and over use a component method in which:

$$HHP65+_{t} = HHP65+_{t-1} + NI65+ + NMIG65+ + NETMOVE65+. \quad (2)$$

NI65+ is natural increase (decrease), which is estimated as the number of persons who aged from 64 to 65 during the year (AGE in equation(1)) minus deaths in the population aged 65 and over.

NMIG65+ is internal migration, which is estimated from Medicare enrollment data. A migration rate is estimated as [(actual Medicare enrollment$_{t-1}$ − expected Medicare enrollment) / actual enrollment$_{t-1}$]. Expected Medicare enrollment is [actual enrollment$_{t-1}$ + (NI65+$_{t-1}$ x the 1990 Medicare coverage ratio)].[4] The estimated migration rate is then applied to the migration base, HHP65+$_{t-1}$ + 0.5(NI$_{t-1}$ + NETMOVE$_{t-1}$).

NETMOVE65+ is other net movements (legal immigrants, undocumented immigrants, refugees, emigrants, net entrants from Puerto Rico), which are estimated as described above for the household population under age 65.

Group Quarters Population Under Age 65 and Age 65 and Over Group quarters populations for both age groups (under age 65 and age 65 and over) are estimated as the 1990 census group quarters population plus the difference between the current group quarters report (GQR) minus the 1990 GQR figure. The GQR is compiled annually from data obtained from the FSCPE, DoD, Veterans Administration, and colleges by type of group quarters: correctional facility, juvenile facility, nursing home, other institutional, college, military quarters, and other noninstitutional.

County Estimates by Age

County age estimates are prepared in a two-step procedure. In the first step, estimates of total county population are developed as described above. Separately, estimates of state populations by single years of age, sex, race, and Hispanic origin are developed. The state age estimates (which are controlled to the state total population estimates) use a component method in which migration rates by age for people under age 65 are derived from school enrollment data (U.S. Census Bureau, 1987).[5]

In the second step, the county age estimates are developed by using a raking-ratio adjustment of the numbers from the previous census. In this approach, the

[4]Previously, the method simply used the change in Medicare enrollment to estimate the migration rate for the population aged 65 and over directly; the current method preserves the county variation in Medicare coverage.

[5]Recently, the Census Bureau developed experimental state estimates of the population by age, sex, race, and Hispanic origin by a cohort-component method in which federal income tax returns are used to estimate net migration on the basis of estimates of gross inmigration and gross outmigration (see National Research Council, 2000:Ch.3). For this experimental method, the resulting state age-sex-race-Hispanic origin estimates are controlled to the state age-sex population estimates developed as described in the text.

beginning matrix of counts for each county by age, sex, race, and Hispanic origin from the previous census is simultaneously adjusted to agree with the postcensal estimate of the total county population and the postcensal estimates for the applicable state by age, sex, race, and Hispanic origin. Beginning with the revised county age estimates for 1994, this adjustment is carried out separately for persons in each age group who were in group quarters in the census and persons who were not in group quarters.

The raking-ratio procedure used for county age estimates assumes that the age distribution of each county within a state changes in the same manner as that state's age distribution. Errors in the county estimates of an age group can arise from errors in this assumption, errors in the derivation of the state estimates of age groups, and errors in the derivation of the county estimates of total population.

School District Population Estimates

The Census Bureau uses a shares approach, similar to that used for distributing the number of poor school-age children among the school districts in a county, to estimate the total population and total school-age population for school districts. The method for producing 1996 estimates of total population and total school-age children for districts involved the following steps: retabulate the 1990 census data according to 1995-1996 school district boundaries, determine the 1990 census county share in each district or part of a district for total population and total school-age children, and apply those shares to the Census Bureau's 1996 county estimates of total population and total school-age children, respectively, derived by the procedures described above. Unlike the situation with poor school-age children, the 1990 census school district shares for total population and school-age population are based on data from the complete count (short form) and are not subject to sampling error.

EVALUATION OF COUNTY ESTIMATES

The Census Bureau has an active program to develop and review the performance of its demographically based state and county population estimates, including evaluating the estimates at 10-year intervals by comparing them with the decennial census. These comparisons provide an indication of the differences, but they are not perfect measures of accuracy and precision because the standard (i.e., the decennial census) itself is flawed, notably from net population undercount, which varies by age group across time and place (see Robinson et al., 1993).

The Census Bureau's methods and data for producing postcensal population estimates have generally improved over time, but three patterns of differences, which are practically inevitable, continue to affect the state and county estimates (see Davis, 1994). First, the proportional differences of the estimates in comparison with the census are larger on average for small areas than for large ones. Second, the proportional differences tend to be larger for areas in which the population is changing rapidly than for areas that are more stable. Third, the proportional differences for age groups tend to be higher than those for the total population.

Comparisons with 1990 Census County Estimates

The Census Bureau conducted an evaluation of the county estimates of total population and children aged 5-17 by comparison with the 1990 census numbers for all counties and for categories of counties. Updated estimates for counties by age were produced by ratio adjusting the 1980 census county age numbers to 1990 county total population estimates and 1990 state age estimates, as described above. The resulting 1990 county age estimates were compared with the 1990 census county age numbers.

Tables 8-1 to 8-8 show the average proportional algebraic difference and the average proportional absolute difference, expressed as percents, between the 1990 county population estimates for people aged 5-17, developed by raking the 1980 census estimates as described above, and the 1990 census numbers.[6] The two measures are shown for all counties and for counties grouped into categories for the following characteristics: population size in 1990; population growth from 1980 to 1990; percentage of black and other nonwhite population in 1990; percentage of Hispanic population in 1990; percentage of poor population in 1990; percentage of group quarters residents in 1990; census geographic division; and metropolitan status. Also shown is the percentage of counties with negative differences (underpredictions relative to the census).

The overall average proportional absolute difference in the 1990 county estimates of people aged 5-17 is 6.3 percent (shown in Table 8-1). By comparison, for 1990 county estimates of total population, prepared using the Census Bureau's current estimation procedure, it is 3.6 percent (Davis, 1994).[7] The average proportional absolute differences do not seem to be concentrated in any

[6]The average proportional absolute difference is computed as the sum for all counties n (or all counties in a category) of the absolute difference between the estimate and the 1990 census figure for each county as a proportion of the census figure for each county, divided by the number of counties, or $\Sigma\ [(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}\ ] / n$. The average proportional algebraic difference is computed similarly, except that the sign of the difference (positive or negative) is considered in the computation.

[7]The average absolute differences for 1990 county estimates of children aged 5-17 and the total population are smaller than the average proportional absolute differences–the average absolute differences are 4.9 percent and 2.3 percent of the county average school-age and total population, respectively (see Tables 8-9 and 8-10).

TABLE 8-1 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Population Size in 1990

Population Size, 1990	Counties (Number)[a]	Average Proportional Algebraic Difference[b]		Average Proportional Absolute Difference[b]		Percentage of Counties with Negative Differences
All	3,140	–0.4	(8.7)	6.3	(6.1)	56.2
1,000,000 and over	30	1.5	(6.5)	5.2	(4.1)	46.7
500,000 to 1,000,000	67	1.7	(5.1)	4.4	(3.2)	29.9
100,000 to 500,000	361	0.9	(5.9)	4.6	(3.8)	50.4
50,000 to 100,000	384	0.6	(7.3)	5.6	(4.7)	51.3
10,000 to 50,000	1,543	–0.5	(7.7)	6.0	(4.8)	56.9
5,000 to 10,000	457	–1.5	(9.0)	7.2	(5.6)	61.7
2,500 to 5,000	180	–3.2	(10.5)	8.4	(7.0)	67.2
Less than 2,500	118	0.0	(21.2)	12.4	(17.2)	59.0

[a]Excludes Kalawao County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Differences are in percent. See text for formulas. Standard deviations are in parentheses.

SOURCE: Data from U.S. Census Bureau.

TABLE 8-2 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Growth Rate, 1980-1990

Population Growth Rate, 1980-1990	Counties (Number)[a]	Average Proportional Algebraic Difference[b]		Average Proportional Absolute Difference[b]		Percentage of Counties with Negative Differences
All	3,140	–0.4	(8.7)	6.3	(6.1)	56.2
Decrease of 5% or more	834	–0.7	(10.1)	6.7	(7.5)	58.4
–5% to 0%	595	–1.0	(7.8)	5.8	(5.3)	60.7
0 to 5%	583	–0.4	(7.8)	5.9	(5.0)	55.8
5 to 10%	386	–0.1	(7.7)	5.7	(5.1)	57.5
10 to 15%	208	0.6	(7.5)	6.1	(4.5)	49.0
15 to 25%	247	0.3	(9.0)	6.7	(6.0)	51.8
25% and over	287	–0.2	(10.0)	7.5	(6.5)	48.4

[a]Excludes Kalawao County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Differences are in percent. See text for formulas. Standard deviations are in parentheses.

SOURCE: Data from U.S. Census Bureau.

TABLE 8-3 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Percent Black and Other Nonwhite Population, 1990

Percent Black and Other Nonwhite, 1990	Counties (Number)[a]	Average Proportional Algebraic Difference[b]		Average Proportional Absolute Difference[b]		Percentage of Counties with Negative Differences
All	3,140	–0.4	(8.7)	6.3	(6.1)	56.2
Less than 0.5%	304	–0.7	(13.1)	7.4	(10.8)	61.6
0.5 to 1.0%	405	–2.1	(8.3)	6.1	(6.0)	67.2
1.0 to 2.0%	468	–1.6	(8.4)	6.8	(5.2)	62.4
2.0 to 5.0%	550	–1.0	(7.8)	6.2	(4.8)	58.6
5.0 to 15.0%	641	0.3	(8.2)	6.3	(5.3)	49.1
15.0 to 40.0%	546	0.8	(7.6)	5.7	(5.1)	48.7
40.0% and over	226	1.6	(8.0)	6.1	(5.5)	48.5

[a]Excludes Kalawao County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Differences are in percent. See text for formulas. Standard deviations are in parentheses.

SOURCE: Data from U.S. Census Bureau.

TABLE 8-4 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Percent Hispanic Population, 1990

Percent Hispanic, 1990	Counties (Number)[a]	Average Proportional Algebraic Difference[b]		Average Proportional Absolute Difference[b]		Percentage of Counties with Negative Differences
All	3,140	–0.4	(8.7)	6.3	(6.1)	56.2
Less than 0.5%	983	0.7	(9.2)	6.1	(6.9)	52.9
0.5 to 1.0%	760	0.2	(7.0)	5.5	(4.3)	52.5
1.0 to 2.0%	485	–0.1	(8.3)	6.5	(5.2)	56.1
2.0 to 5.0%	385	–1.4	(8.8)	6.5	(6.2)	60.3
5.0 to 15.0%	291	–3.2	(9.8)	7.5	(7.0)	63.4
15.0 to 40.0%	162	–3.8	(10.4)	8.4	(7.3)	70.4
40.0% and over	74	–1.0	(8.1)	6.5	(4.9)	56.8

[a]Excludes Kalawao County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Differences are in percent. See text for formulas. Standard deviations are in parentheses.

SOURCE: Data from U.S. Census Bureau.

TABLE 8-5 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Percent Poor Population, 1990

Percent Poor, 1990	Counties (Number)[a]	Average Proportional Algebraic Difference[b]		Average Proportional Absolute Difference[b]		Percentage of Counties with Negative Differences
All	3,140	–0.4	(8.7)	6.3	(6.1)	56.2
None	50	–1.5	(8.2)	7.0	(4.4)	58.0
Less than 5%	253	1.8	(13.9)	7.3	(11.9)	45.5
5 to 10%	1,046	–1.4	(7.7)	6.0	(5.1)	62.1
10 to 15%	929	–1.1	(8.4)	6.7	(5.2)	58.3
15 to 25%	688	0.9	(8.2)	6.1	(5.6)	50.0
25 to 40%	157	0.8	(7.0)	5.4	(4.4)	49.7
40% and over	17	3.1	(11.4)	8.9	(7.5)	35.3

[a]Excludes Kalawao County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Differences are in percent. See text for formulas. Standard deviations are in parentheses.

SOURCE: Data from U.S. Census Bureau.

TABLE 8-6 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Percent Group Quarters Residents, 1990

Percent Group Quarters, 1990	Counties (Number)[a]	Average Proportional Algebraic Difference[b]		Average Proportional Absolute Difference[b]		Percentage of Counties with Negative Differences
All	3,140	–0.4	(8.7)	6.3	(6.1)	56.2
Less than 0.5%	175	1.0	(17.3)	9.9	(14.2)	50.6
0.5 to 1.0%	372	1.4	(8.1)	6.4	(5.2)	43.0
1.0 to 1.5%	636	0.6	(7.6)	5.9	(4.7)	49.7
1.5 to 2.0%	591	–0.3	(7.4)	5.7	(4.6)	55.3
2.5 to 3.0%	535	–1.8	(8.3)	6.3	(5.7)	64.7
3.0 to 5.0%	431	–0.9	(7.0)	5.5	(4.4)	60.8
5.0% and over	400	–2.1	(9.0)	7.1	(5.9)	66.1

[a]Excludes Kalawao County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Differences are in percent. See text for formulas. Standard deviations are in parentheses.

SOURCE: Data from U.S. Census Bureau.

TABLE 8-7 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Census Division

Census Division	Counties (Number)[a]	Average Proportional Algebraic Difference[b]		Average Proportional Absolute Difference[b]		Percentage of Counties with Negative Differences
All	3,140	–0.4	(8.7)	6.3	(6.1)	56.2
New England	67	–1.2	(5.3)	4.1	(3.6)	62.7
Middle Atlantic	150	0.6	(5.2)	4.1	(3.3)	54.0
East North Central	437	–1.4	(5.7)	4.7	(3.5)	64.5
West North Central	618	–3.0	(7.5)	6.4	(5.0)	72.0
South Atlantic	591	2.4	(8.3)	6.5	(5.7)	39.6
East South Central	364	2.9	(7.2)	6.0	(5.0)	37.4
West South Central	470	–0.4	(9.6)	7.0	(6.5)	50.9
Mountain	281	–2.8	(14.1)	9.0	(11.2)	68.7
Pacific	161	–2.5	(7.9)	6.5	(5.1)	68.5

[a]Excludes Kalawao County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Differences are in percent. See text for formulas. Standard deviations are in parentheses.

SOURCE: Data from U.S. Census Bureau.

TABLE 8-8 Evaluation of 1990 County Population Estimates for Age Group 5-17, by Metropolitan Status, 1990

Metropolitan Status, 1990	Counties (Number)[a]	Average Proportional Algebraic Difference[b]		Average Proportional Absolute Difference[b]		Percentage of Counties with Negative Differences
All	3,140	–0.4	(8.7)	6.3	(6.1)	56.2
Nonmetropolitan	2,393	–1.2	(9.0)	6.5	(6.3)	60.0
Metropolitan	747	1.9	(7.2)	5.6	(4.9)	43.9

[a]Excludes Kalawao County, Hawaii, which had no persons aged 5-17 in 1980 or 1990.
[b]Differences are in percent. See text for formulas. Standard deviations are in parentheses.

SOURCE: Data from U.S. Census Bureau.

particular types of counties (Tables 8-1 to 8-8), except that, as one would expect, the smallest counties (those with populations under 2,500) have differences running at twice the overall average: 12.4 percent, compared with 6.3 percent overall (see Table 8-1).

There may be a systematic prediction bias by county population size (Table 8-1). The average proportional algebraic difference is negative (indicating underestimates) for counties in the smaller population size groups (except for those under 2,500 with a 0.0 value) and positive (indicating overestimates) for counties in the larger population size groups. The percentage of counties with negative differences generally increases as county population size decreases. Nonmetropolitan counties also have a negative average proportional algebraic difference (see Table 8-8), with 60 percent of these counties having negative differences, which is consistent with the pattern of negative differences for smaller counties. Negative average proportional algebraic differences also characterize counties with negative or lower rates of population growth (Table 8-2); with lower percentages of black and other nonwhite population (Table 8-3); with average or higher than average percentages of Hispanic population (Table 8-4); with smaller percentages of poor population (Table 8-5); with higher percentages of group quarters residents (Table 8-6); and for counties in the Mountain, Pacific, North Central (East and West), and New England Divisions (Table 8-7).

An issue in examining the average proportional algebraic differences in the 1990 county estimates of children aged 5-17 for categories of counties is whether the patterns observed—for example, the tendency for smaller (larger)-sized counties to have negative (positive) differences—are statistically significant, suggesting the possibility of a systematic bias. Tests of significance were conducted to determine whether there is evidence of possible bias with respect to the characteristics in Tables 8-1 to 8-8.[8]

The test results suggest the possibility of some bias associated with the estimates of children aged 5-17 for several categories of counties: county population size, percentage of black and other nonwhite population, percentage of Hispanic population, percentage of group quarters residents, metropolitan status, and census geographic division. However, the results are not conclusive given that there is only a single year—1990—for which it is possible to evaluate the estimates by comparison with figures from the census or another source.

[8]Since most of these characteristics have ordered categories, a test of a linear trend was conducted using the Abelson-Tukey test procedure (Abelson and Tukey, 1963). Because the number of degrees of freedom is large, the test statistic has essentially a normal distribution under the null hypothesis of no trend. The categories for census geographic division do not have an ordering, so a one-way analysis of variance was performed for that characteristic.

Effects on Poverty Estimates

An issue in the context of Title I allocations is the extent to which errors in the population estimates for children aged 5-17 affect the estimates of the proportion of poor school-age children from log number models, or how they affect the estimates of the number of poor school-age children from log rate models. In the aggregate, the use of population estimates to convert estimated numbers from log number models to estimated proportions added about 1 percentage point to the overall average proportional absolute difference between the model estimates for 1989 and the 1990 census estimates (compare column 3 with column 2 of Table 6-3 in Chapter 6 for the two log number models). The use of population estimates to convert estimated proportions from log rate models to estimated numbers had even less effect overall (compare column 2 with column 3 of Table 6-3 for the two log rate models).

In addition, although a rigorous analysis was not done, there seems to be little systematic contribution of errors in the population estimates to category differences in the model estimates of poor school-age children from the 1990 census estimates (see Appendix C). For the three single-equation rate models that were examined for 1989 in the first round of evaluations–the log rate model (under 21), the rate model, and the hybrid rate-number model (see Chapter 5)–the use of population estimates instead of 1990 census numbers to convert estimated proportions to estimated numbers of poor school-age children worsened the performance of the models for some characteristics (e.g., by increasing the spread between the largest negative and positive category differences compared with the census), improved their performance for other characteristics, and made essentially no difference for other characteristics. None of the category differences between the model estimates of poor school-age children developed with population estimates and those developed with 1990 census numbers was large.

The evaluations of the effects of the population estimates on estimates of poor school-age children outlined above relate to a 10-year period: the population estimates for 1990 were developed on the basis of 1980 census data updated with other sources. The 1994 population estimates that are used to convert estimated numbers to estimated proportions of poor school-age children in 1993 from the log number (under 18) model were developed on the basis of 1990 census data. Because of the 4-year instead of 10-year period for updating, it is likely that errors in the 1994 population estimates are smaller than errors in the 1990 population estimates and that they have even smaller effects on the estimates of the number and proportion of poor school-age children. Errors in the 1996 population estimates that are used to convert estimated numbers to estimated proportions of poor school-age children in 1995 may also be somewhat smaller than errors in the 1990 population estimates.

EVALUATION OF SCHOOL DISTRICT ESTIMATES

As it did for the school district estimates of poor school-age children, the Census Bureau evaluated its method for estimating total population and total school-age children at the district level by using the 1980-1990 evaluation file (see Chapter 7) to compare three sets of 1990 school district estimates with 1990 census numbers. The three sets of estimates were derived by: (1) applying 1980 census school district shares within counties to 1990 county population estimates; (2) applying 1980 census school district shares within counties to 1990 census county numbers; and (3) applying 1980 census school district shares within the nation as a whole to the national 1990 census number.

Tables 8-9 and 8-10 provide summary statistics for the three sets of school district estimates of 1990 total population and 1990 total school-age children, respectively, compared with the 1990 census numbers. The statistics provided are the average absolute difference between the estimates from a method and the census expressed as a percent of the average total population or total school-age children in the census, and the average proportional absolute difference between each set of estimates and the 1990 census numbers. For comparison purposes, the last row of each table provides the same statistics for county estimates of total population and total school-age children in 1990 from the Census Bureau's population estimates program. (As noted above, this program uses administrative records, such as births and deaths, to update population numbers from the previous census.)

The county estimates of total population and total school-age children for 1990 differ little from the 1990 census numbers: the average absolute differences are 2 percent and 5 percent, respectively (Tables 8-9 and 8-10, first column); the average proportional absolute differences are 4 percent and 6 percent, respectively. The school district estimates show larger differences, although the differences are much smaller than those for school district estimates of poor school-age children (see Table 7-3 in the previous chapter). For school district estimates of total population under method (1), the average absolute difference is 10 percent of the average total population; for school district estimates of total school-age children under method (1), the average absolute difference is 12 percent of the average total school-age children. By comparison, for school district estimates of poor school-age children under method (1), the average absolute difference is 22 percent of the average number of poor school-age children. The corresponding average proportional absolute differences are 13 percent (total population), 17 percent (total school-age children), and 60 percent (poor school-age children).

As noted above, evaluations of Census Bureau population estimates for states and counties have shown that the proportional differences of the estimates in comparison with census numbers are larger on average for small areas than for large ones. The proportional differences of the estimates also tend to be larger for areas in which the population is changing rapidly than for areas that are more

TABLE 8-9 Comparison of Within-County Shares Estimates and 1990 Census School District Numbers of Total Population in 1990

Model	Average Absolute Difference, Relative to Average Total Population (in percent)[a]	Average Proportional Absolute Difference (in percent)[b]
1990 School District Estimates		
(1) Within-county shares method using 1980 census shares applied to 1990 county model estimates	9.6	13.3
(2) Within-county shares method using 1980 census shares applied to 1990 census county numbers	9.2	12.6
(3) National stable shares method using 1980 census shares applied to 1990 census national number	13.9	18.9
1990 County Estimates from Census Bureau's Population Estimates Program	2.3	3.6

NOTES: School district estimates are based on 9,201 districts (9,243 districts in the 1980-1990 evaluation file minus 42 districts with estimated population 30 or less in 1980 or 1990). The 1990 census numbers used in the comparisons are from the complete count and are not subject to sampling error. The estimates from the three methods are controlled to the 1990 census national total population number before comparison to the 1990 census school district estimates.

[a]The formula, where there are n school districts or counties (i), and Y is the estimate (number) for the total population from a model (census), is

$$\Sigma[(|Y_{\text{model } i} - Y_{\text{census } i}|) / n] / [\Sigma(Y_{\text{census } i}) / n] .$$

[b]The formula is $\Sigma [(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}] / n$.

SOURCE: Data from U.S. Census Bureau.

stable. The school district estimates of total population and total school-age children follow the same patterns.

Compared with the school district estimates of poor school-age children, the estimates of total population and total school-age children benefit from two factors. First, total population and total school-age children are larger quantities to estimate. Second, the census data that are used to form within-county school district shares of total population and total school-age children, while subject to measurement error, are obtained from a complete count. Nonetheless, the estimates of total population and total school-age children for school districts are not nearly as accurate as the corresponding county estimates. The Census Bureau has

TABLE 8-10 Comparison of Within-County Shares Estimates and 1990 Census School District Numbers of Total School-Age Children in 1990

Model	Average Absolute Difference, Relative to Average Total School-Age Children (in percent)[a]	Average Proportional Absolute Difference (in percent)[b]
1990 School District Estimates		
(1) Within-county shares method using 1980 census shares applied to 1990 county model estimates	12.0	16.9
(2) Within-county shares method using 1980 census shares applied to 1990 census county numbers	10.4	16.1
(3) National stable shares method using 1980 census shares applied to 1990 census national number	16.6	20.6
1990 County Estimates from Census Bureau's Population Estimates Program	4.9	6.3

NOTES: School district estimates are based on 9,201 districts (9,243 districts in the 1980-1990 evaluation file minus 42 districts with estimated population 30 or less in 1980 or 1990). The 1990 census numbers used in the comparisons are from the complete count and are not subject to sampling error. The estimates from the three methods are controlled to the 1990 census national number of total school-age children before comparison to the 1990 census school district estimates.

[a]The formula, where there are n school districts or counties (i), and Y is the estimate (number) of total school-age children from a model (census), is

$$\Sigma[(|Y_{\text{model } i} - Y_{\text{census } i}|) / n] / [\Sigma(Y_{\text{census } i}) / n] .$$

[b]The formula is $\Sigma [(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}] / n$.

SOURCE: Data from U.S. Census Bureau.

begun, but has not had time to complete, an analysis of school enrollment data to determine if these data could be used to improve the school district estimates of total school-age children. Such work should be continued (see Chapter 9; see also National Research Council, 2000:Chapter 5).

9

Research and Development Priorities

There are several reasons that make it important for the Census Bureau to continue an active program of research and development for methods of estimating poverty for school-age children at the county and school district levels. For counties, although there is clear evidence that the county model is performing reasonably well, the county (and state) model evaluations have identified a number of issues that warrant investigation as a priority in the short term to determine how to further improve the estimation procedures. Also, with a model-based approach, it is important to examine carefully the continued applicability of a model each time it is used and to modify it appropriately when necessary. In addition, research is needed to take account of likely future developments in the availability and characteristics of data sources that have implications for the modeling effort and to work on longer term modeling issues. Continued work to improve the county model is important not only for county estimates, but also to improve school district estimates that are developed by using the within-county shares estimation procedure.

For school districts, the important short-term priority is to investigate ways to improve the within-county shares method for developing updated estimates of total and poor school-age children. Also, it is not too soon to begin research on ways to take advantage of likely future developments in available data that could make it possible to develop an estimation method that (unlike the shares method) captures changes in school-age poverty among districts within counties as well as changes between counties.

This chapter identifies short-term priorities for research and development of the current Census Bureau models for estimates of poor school-age children. The panel's final report (National Research Council, 2000) discusses these short-term priorities as well. It also develops an agenda for longer-term work to take advantage of possible new sources of survey and administrative records data that could improve the Bureau's models for poor school-age children and the other income and poverty estimates that are produced by the Bureau's SAIPE Program.[1]

The chapter begins by reviewing the schedule for the Census Bureau to provide updated small-area estimates of poor school-age children. It then considers short-term research priorities for county and school district estimates and mentions some longer term priorities as well. It concludes by noting the requirements for an ongoing program of small-area income and poverty estimates, particularly for thorough evaluation and full documentation of models and results (see also National Research Council, 2000:Ch.7).

SCHEDULE CONSIDERATIONS

The next three legislatively mandated deadlines for the Census Bureau to deliver updated school district estimates of poor school-age children to the Department of Education for use in Title I allocations are as follows:

- October 2000: estimates for 1997 (or later) for use for allocations for the 2001-2002 and 2002-2003 school years
- October 2002: estimates for 1999 (or later) for use for allocations for the 2003-2004 and 2004-2005 school years
- October 2004: estimates for 2001 (or later) for use for allocations for the 2005-2006 and 2006-2007 school years

In each case, three estimates are needed for each school district: numbers of poor and of total school-age children and the total population. Although the legislation does not require county estimates, they will be needed as long as the method for producing school district estimates includes an adjustment or control to county estimates. There is also interest in state and county estimates of poor children for other important public policy uses, such as evaluating the effects of changes in welfare programs.

Priorities for short-term and longer term research should consider the impor-

[1]The final report considers the possible role for the SAIPE Program of two new survey data sources with relatively large sample sizes, the 2000 census long-form survey and the planned American Community Survey, as well as two smaller ongoing surveys, the March CPS and the Survey of Income and Program Participation. The report also considers the role of improvements to the Census Bureau's Master Address File and associated geographic coding system in making it possible to use administrative records to develop poverty estimates for school districts and other subcounty areas.

tant changes that are likely to occur in the availability of data for modeling over the next 5 years and beyond, which include:

- current and future changes to welfare programs and tax systems that may affect the comparability or applicability of Food Stamp Program and Internal Revenue Service (IRS) data for use in small-area estimation models;
- the income and poverty estimates for small areas that will be available from the 2000 decennial census long-form sample of about 18 million households beginning in 2002; and
- the planned introduction of the American Community Survey (ACS) as a large-scale, continuing sample survey of U.S. households, conducted primarily by mail, that will provide estimates similar to those provided by the decennial census long-form sample, including income and poverty estimates for small areas. The ACS is currently under development. Beginning in 2003, the full ACS sample will be 250,000 housing units each month throughout the decade, for an annual sample size of about 3 million housing units spread across all counties in the nation. The current plan is that the ACS, like the 2000 census long form, will oversample small jurisdictions. Unlike the 1990 census, the oversampling in the 2000 census and the ACS includes small school districts.[2]

SHORT-TERM PRIORITIES

County Estimates

In its third interim report (National Research Council, 1999), the panel identified seven types of research that should be pursued as a priority to determine if the current estimation procedure for counties can be improved: modeling of CPS county sampling variances; estimation of model error and sampling error variance in the state model; methods to incorporate state effects in the county model; discrete variable models that include counties in the CPS sample that have no sampled households with poor school-age children; ways to reduce the time lag of the estimates; evaluation of food stamp and other input data; and large category differences and residual patterns for the state and county models. Since then, the Census Bureau has made progress in several of these research areas as noted below.

Modeling of CPS County Sampling Variances The residual variance for the county model comprises two components: the model error variance and the sampling error variance of the dependent variable. These two components need

[2]For information about the ACS, see Alexander (1998); the Census Bureau's web site: http://www.census.gov/acs/www; and National Research Council (2000:Ch.4, 5).

to be reasonably well estimated for the application of the model (e.g., to determine the relative weights of the regression estimate and the direct estimate in the shrinkage procedure). The current approach for estimating these components is to assume that the model error variance from the 1989 regression equation with the dependent variable formed from 1990 census data is the same as the model error variance when the dependent variable is formed from the 3 years of CPS data that are used for the county model equation for the estimation year. The total sampling error variance is then obtained simultaneously with the regression parameter estimates through use of maximum likelihood estimation. As part of this procedure, the sampling variance for a particular county is assumed to be inversely proportional to the CPS sample size in that county.

There is ample evidence that the function that is now used to distribute the total sampling variance to counties is incorrect (see Chapter 6). The Census Bureau's experimentation with other functions (specifically, investigating a function in which the sampling variance is inversely proportional to the square root of the CPS sample size in a county–see Fisher and Asher, 1999a) should be pursued to eliminate or reduce the problem of variance heterogeneity with respect to both the CPS sample size and the predicted value of the number of poor school-age children that is evident in the county model regression output. Research on this topic should include an assessment of the effects of alternative variance functions on the county estimates.

In addition, the Census Bureau should continue to pursue an alternative approach, which is to estimate the CPS sampling variances for counties with adequate sample size on the basis of direct calculations of these variances that take account of the clustered sample design within these counties, and then use a generalized variance function for modeling the sampling variances for all counties with CPS-sampled households. With this approach, the model error variance is then obtained simultaneously with the regression parameter estimates through use of maximum likelihood estimation, as in the state model. The Census Bureau's work on fitting a generalized variance function to the CPS sampling variances should continue and should include an assessment of the effects on the county estimates to determine if the benefits justify continued refinement of the variance modeling.

Model Error and Sampling Error Variance in the State Model In the state model the model error variance is obtained from a maximum likelihood procedure that estimates the coefficients of the predictor variables and the model error variance, given estimates of the sampling error variances of the direct state estimates. For most years for which the state model has been estimated, this procedure estimates the model error variance as zero, which results in zero weight being given to the direct CPS estimates. In effect, the model is assumed to be without error, which is not credible. A likely explanation is that the Census Bureau's estimates of sampling error variance for the direct state estimates are overestimates, which results in a value of zero for the model error variance when the state sampling variances are used in a maximum likelihood procedure that estimates the coefficients of the predictor variables and the model error variance. The Census Bureau should continue to investigate its procedures for estimating sampling error variance. It should also examine the effects of a simple correction, such as putting a small weight on the direct estimates in weighting the estimates from the CPS equation for a target year.[3]

State Effects The magnitude of the state raking factors that are used to adjust the county estimates warrants further investigation. Preliminary calculations by the panel suggest that sampling error may account for much, but not all, of the variation in the raking factors. The Census Bureau should conduct further research to better understand the causes of this variation. One part of this research could be to examine the effect of using 3 years rather than 1 year of CPS data in the state model, as is done in the county model.

More generally, work should be conducted to determine if there are idiosyncratic state effects that should be captured in the county model. The Census Bureau did some preliminary research on adding fixed state effects to alternative formulations of the county model (see Appendix A). While the addition of fixed state effects reduced some nonrandom residual patterns in the regression output, a fixed state effects model did not perform better than other models in comparison with the 1990 census estimates (see Appendixes B and C). Some preliminary work with a random state effects model with two components of variance, one for state and one for county within state (see Fuller and Goyeneche, 1998), suggested that a small state random effect may be present and that further research on a random state effects model should be conducted.

Discrete Variable Models that Use Counties with No Sampled Poor School-Age Children When using a logarithmic transformation of the number of poor school-age children as the dependent variable in the county regression model, all counties in the CPS sample for which none of the sampled households has school-age children who are poor (262 of 1,247 counties for the 1995 model) have to be removed from the regression analysis. The dropped counties are generally smaller counties with small CPS sample sizes.

While the dropped counties would have little influence in any regression equation due to their small sample sizes, the exclusion of 21 percent of the counties in the CPS sample is a cause for concern. Moreover, the internal and

[3]Bell (1999) has explored yet another approach, which is to use a Bayesian model to account for the uncertainty in the estimates of the model error variance. This approach yields positive estimates of model error variance that could be useful for producing the state model estimates.

external evaluations of the county model suggest that although the current approach provides reasonably good estimates for small counties for 1989, 1993, and 1995, they could be improved. For example, there is a slight tendency in the county model equation to overpredict poverty in small counties (see Chapter 6). It is important to investigate the development of discrete variable regression models, such as Poisson regression or other forms of generalized linear models, that permit the inclusion of data for those counties that have no sampled families with children in poverty. The Census Bureau has begun work on a hierarchical Bayesian modeling approach that addresses this problem, and this work should continue (see Fisher and Asher, 1999b).

Ways to Reduce the Time Lag of the Estimates The Title I fund allocations for the 1999-2000 school year were based on estimates of school-age children in 1996 who were in poor families in 1995, and these estimates were also used for the 2000-2001 school year allocations. It is important to explore the extent to which this time lag can be reduced for the county estimates, which will correspondingly reduce the time lag for the school district estimates.[4] The Census Bureau began some exploratory work on this topic in June 1997 but had to put it aside. Now that the county estimation procedure has been developed and put on a production basis, it is important to resume this work.

One of the causes of the lag is the availability of food stamp data for counties, which must be obtained from individual states in some instances and which are not available until almost 2 years after the year to which they refer. It might be possible to overcome this problem, without seriously harming the performance of the county model, by using food stamp data for the year prior to the estimation year. Another possibility is to control the estimates from the county model to the state model estimates for the latest of the 3 years of CPS data used in the county model, instead of to the middle year. These ideas and others (see National Research Council, 2000:Ch.3) need to be evaluated to determine if the lag between the time period of the estimates and the year of allocation of funds can be reduced.

Evaluation of Food Stamp and Other Input Data Regular evaluation of the continued suitability of food stamp and other data for input to the state and county models is important for the Census Bureau's small-area estimation program. Changes in welfare programs and the accompanying data systems (especially those resulting from the 1996 Personal Responsibility and Work Opportunity Reconciliation Act) will almost certainly affect the comparability of food stamp

[4] It would also be desirable to reduce the time lag in the school district boundary survey so that the allocations are made to current school districts. However, that survey is conducted every 2 years, and it may not be possible to carry it out more frequently or to complete it more quickly.

data over geographic areas. For example, legal immigrants, many of whom are no longer eligible for benefits, are very unevenly distributed geographically. Comparability is an important assumption in both the county and state regression models, and, therefore, the way in which food stamp data are used as a predictor variable in the models may need to be modified. Changes in the tax system could also affect the usefulness of IRS data for small-area poverty estimation. More generally, it is important to continually evaluate the input data to the state and county models to assess errors or inconsistencies in them and to develop methods to account for those errors in the modeling process.

Large Category Differences and Residual Patterns for the State and County Models The internal and external evaluations (see Chapter 6) demonstrated that the state and county models are generally well behaved with respect to the estimates for various categories of states and counties. However, it is important to investigate further the residual patterns and category differences to determine if the regression models could be improved either through a modification of the model form or through the addition of predictor variables.[5]

As an example of a pattern that is worth further investigation, when compared with CPS aggregate estimates, the county model exhibited a tendency in 1989, 1993, and 1995 to underpredict the number of poor school-age children in counties with large percentages of Hispanics. Also, from examination of the standardized residuals, the state model exhibited a tendency to underpredict the proportion of poor school-age children in some states in the West Region.

More generally, as a model is estimated for additional years, it is important to look for consistent patterns of residuals and category differences to understand their causes and to take corrective action when necessary. While it may be necessary to tolerate overprediction or underprediction for a particular type of area in any one year, a consistent pattern of overprediction or underprediction needs to be addressed.

In the evaluation of residuals and category differences, particular attention should be paid to states and counties that have experienced large demographic or socioeconomic changes that may correlate with changes in numbers of poor

[5]The evaluations conducted to date of the county estimates include examination of the residual patterns from the regression model, comparisons of the model estimates for 1989 with 1990 census estimates, and comparisons of the model estimates for 1989, 1993, and 1995 with aggregate CPS estimates. Another evaluation that could help determine what portion of the errors in the county estimates is due to problems with the model–rather than measurement differences and sampling variability–is to fit the model to 1990 census data (prior to shrinkage and raking to the state model) and to compare the estimates to 1990 census values for aggregates of counties. This evaluation is similar to the county model-CPS aggregate comparisons, but it has the advantage that the sampling error in the census is much less than in the CPS. The county model estimates are not shrunk for this evaluation because the resulting estimates would have considerable weight on the census direct estimates and so be less informative about possible problems with the regression model.

school-age children. For example, the federal tax return data that are used to estimate internal migration for the postcensal population estimates might be used to classify states and counties into categories by migration rates and the performance of the models compared for these categories. Also, the performance of the models might be compared for categories of counties classified by overall population change since the 1990 census. In turn, adding predictor variables to the models from the decennial census and the population estimates program, possibly including interaction terms, may prove a fruitful way to address persistent patterns of overprediction or underprediction for these and other categories of states and counties.

School District Estimates

There cannot be marked improvements in the school district estimates without a substantial effort to improve the data sources for districts and to develop models to use them. Nonetheless, work should go forward to further evaluate the current estimation method and to seek to effect modest improvements in it. Three important areas for research are: investigation of methods to reduce the variance of the census estimates of poor school-age children; use of school enrollment data to improve estimates of the total number of school-age children; and investigation of the possible use of National School Lunch Program data to improve estimates of poor school-age children.

Reducing the Variance of the Census Estimates of Poor School-Age Children Because so many school districts are so small in size, the census estimates of poor school-age children, which derive from the long-form sample, are subject to high sampling variability. In addition to affecting the quality of the 1995 school district estimates that were developed by the Census Bureau's within-county shares method, the sampling variability in the 1990 census estimates affected the 1980-1990 evaluations. The evaluation measures reported in Chapter 7 overstate the degree of error in the within-county shares estimates because of this sampling variability. The Bureau should continue its research in partitioning out the sampling error from the root mean square difference between the within-county shares estimates and the census estimates of poor school-age children (see Chapter 7) in order to produce a better indicator of the quality of the school district estimates.

The 1990 census school district estimates of poor school-age children that were used in the 1995 estimates and as the standard of comparison in the 1980-1990 evaluations were developed by ratio adjustment. This procedure, which applies the long-form-sample-based estimates of the school-age poverty rate to the complete-count estimates of total school-age children, reduces the variance of the 1990 census estimates to a modest extent. Other ways to further reduce the variance should be investigated.

One approach is to incorporate other characteristics from the census short form that are known to be related to poverty in estimating school district numbers of poor school-age children from the census. For example, such characteristics as race and ethnicity, home tenure (owner, renter), family type, and residence (e.g., central city) could be used for this purpose. A very simple form of this type of estimation procedure would be a stratified ratio adjustment with strata defined using short-form information.

Another approach is to smooth the census school district estimates with the census county estimates. By carefully constructing smoothed school-district estimates as combinations of school-district and county-level estimates, it might be possible to produce school-district estimates with lower mean square errors than the direct census estimates. It would be desirable to make use of knowledge about model error and sampling variances at the school-district level—if available—to tailor the degree of smoothing for each school district. If successful, smoothing procedures might substantially improve the estimation of census school-age poverty rates in small school districts. They would add some bias because county poverty rates differ from poverty rates for school districts contained within them, but they could potentially substantially reduce variance, thereby improving mean square error.

The development of a smoothing approach should include a thorough evaluation. As part of that evaluation, it would be useful to compare 1990 census estimates of poor school-age children for school districts with three sets of estimates that differ in the calculation of 1980 census within-county shares that are applied to the 1989 county model estimates: unsmoothed 1980 census within-county shares (as in method (1), see Chapter 7); smoothed 1980 census within-county shares; and 1980 census within-county shares that use the 1980 census county school-age poverty rates for all school districts within each county. The third method represents a complete smoothing of the school district poverty rates within counties.

If one or both methods for reducing the variance of the census school district estimates of poor school-age children (smoothing and using other characteristics in the estimation) are successful, then the revised census estimates should be employed with the within-county shares approach if it is used again in the future. The revised estimates from the 1990 census should also be used as the standard of evaluation for assessing the within-county shares estimates of poor school-age children in 1989.

Use of School Enrollment Data to Improve Estimates of the Total Number of School-Age Children The method for estimating total school-age children is similar to that for estimating poor school-age children, namely, to apply census estimates of school district shares within each county to updated county estimates. The method is more robust for total school-age children (and total population) than for poor school-age children because the numbers being estimated are larger and because the census shares for total school-age children (and total population) are based on complete-count data that are not subject to sampling error. But the within-county shares method still does not capture within-county changes in school district populations that have occurred since the census.

Public school enrollment data are collected annually by the National Center for Education Statistics (NCES) for school districts. The Census Bureau has begun research to determine if these data could be used to update the within-county school district shares of total school-age children. Part of this research is to examine reported school enrollment in the 1980 and 1990 censuses for school districts to determine if the within-county enrollment shares in 1990, or, alternatively, the changes in enrollment from 1980 to 1990, produce estimates of total school-age children that are more accurate for 1990 than the 1980 census-based shares. This work should continue. If it is successful, research would also be needed to evaluate the quality of the NCES enrollment data and to determine if such factors as changes in public versus private school enrollment present a problem for estimation.

If it is determined that the use of enrollment data would improve school district estimates of total school-age children, it will be necessary to modify the estimation procedure for poor school-age children so that the estimates of both groups (total and poor) are consistent. One way to achieve consistency would be to apply census school-age poverty rates for districts to the updated estimates of within-county shares of total school-age children that are developed from enrollment data.

Possible Use of School Lunch Data to Improve Estimates of Poor School-Age Children There are many reasons that school lunch data are not necessarily a good proxy for school-age poverty (see Chapter 7). Moreover, at present, there is no complete, accurate source of school lunch data by school district that is readily available to the Census Bureau. Nonetheless, approval to receive free meals under the National School Lunch Program is an indicator of low income, and it seems worthwhile to pursue for other states the research that the panel undertook for New York and Indiana (see also National Research Council, 2000:Ch.5).

The Census Bureau may be able to work through its state data centers for selected states to obtain school lunch data by district for 1989-1990 to evaluate whether within-county school lunch shares in 1989-1990 produce estimates of poor school-age children in 1989 that are more accurate than those produced from the 1980 census-based shares. Another approach to evaluate is whether a combination of school lunch data and census data would be preferable to using either data source alone. The research should also look at the effects of using school lunch data, solely or in combination with census data, to estimate school-age poverty rates because of the role that rates play in concentration grants. If the results of such research are promising, it would be necessary for the NCES to improve the reporting of participation in the National School Lunch Program that it collects in the Common Core of Data.

DOCUMENTATION AND EVALUATION

The development of small-area estimates of income and poverty is a major effort that includes data acquisition and review, database development, geographic mapping and geocoding of data, methodological research, model development and testing, and documentation and evaluation of procedures and outputs. Since the production of small-area poverty estimates supports a range of important public policies for federal, state, and local governments—including the allocation of funds—it is essential that the Census Bureau have adequate staff and other resources for all components of the estimation program, including evaluation and documentation. It is the responsibility of any agency that produces model-based estimates to conduct a thorough assessment of them, including internal and external evaluations of alternative model formulations.

An integral part of the evaluation effort is the preparation of detailed documentation of the modeling procedures and evaluation results. No small-area estimates should be published without full documentation. Such documentation is needed for analysts both inside and outside the Census Bureau to judge the quality of the estimates and to identify areas for research and development to improve the estimates in future years.

Users of small-area estimates of income and poverty from the Census Bureau's SAIPE Program for fund allocation or other purposes should carefully review the documentation provided by the Bureau to understand the properties of the estimates. Users should also study the effects of using the estimates for allocations (see National Research Council, 2000:Ch.6, 7). The Committee on National Statistics is planning to conduct more work in this area. With the participation of our panel, it held a workshop in spring 2000 on issues in using estimates for fund allocation, and a more intensive study of the interactions of properties of estimates with features of funding formulas began in fall 2000. We believe such an effort can usefully inform both users and producers of small-area estimates.

Appendices

APPENDIX

A
Models for County and State Poverty Estimates

William R. Bell
Statistical Research Division, Bureau of the Census

This appendix reviews the models investigated by the Census Bureau for the 1993 county poverty estimates for children aged 5-17; the state model is also reviewed briefly. The same model forms can be used for poverty statistics for other age groups, with appropriately defined dependent and regression variables.

NOTATION

The following notation is used in the estimation program:

- y_{it} = CPS 5-17 poverty estimate for county i in year t;
- Cen_i = previous census estimate for county i (where necessary, a specific census is distinguished by writing $\mathrm{Cen90}_i$ or $\mathrm{Cen80}_i$);
- Y_{it}, Z_i = "true" quantities estimated by y_{it} and Cen_i (i.e., Z_i is not assumed to be true poverty, since the census could be biased relative to CPS);
- e_{it}, ϵ_i = sampling errors in y_{it} and Cen_i, assumed independent $N(0, v_e/n_{it})$ and $N(0, c_i)$, with c_i and n_{it} known, and v_e a parameter to be estimated;
- n_{it} = CPS sample size (number of households) in county i in year t;
- $\mathbf{x}_{it}$, $\mathbf{x}_{i,89}$ = vectors of a constant term (i.e., 1) and regression variables from administrative records for county i in income years t and 1989, respectively;
- β, η = corresponding vectors of regression parameters.

The CPS data that are modeled are for income year (t) 1993 or 1989 (for CPS samples taken in March 1994 and March 1990, respectively). The census data modeled are from the 1990 census and are for income year 1989. The 1980 census data (for income year 1979) enter SAIPE models as regression variables in the equation for the 1990 census data but are not themselves the dependent variable in any model (because the corresponding regression variables $\mathbf{x}_{i,79}$ are not available.)

Note that $y_{it} = Y_{it} + e_{it}$ and $\text{Cen}_i = Z_i + \in_i$. The nature of Y_{it} and Z_i, and their estimators, y_{it} and Cen_i, varies. They can be log(numbers of poor), log(poverty rates), or unlogged poverty rates, depending on the model. Similarly, $\mathbf{x}_{it}$ and $\mathbf{x}_{i,89}$ vary over models. These variations are noted below for the specific models.

The CPS estimates y_{it} and sample sizes n_{it} are 3-year "averages" of CPS estimates centered on year t. The specific formulation depends on whether log(numbers of poor children) are being modeled, as opposed to either child poverty rates or their logarithms (see below for details). Given that y_{it} involves a 3-year average, the corresponding "sample size" n_{it} is defined by counting the number of households in sample in county i in each year of the average ($t - 1$, t, $t + 1$) and adding the three numbers together. For counties with a CPS sample in only 2 of the 3 years, y_{it} is defined from just a 2-year average, and the corresponding n_{it} is defined by summing the households in sample for the 2 years. For counties with a sample in just one of the years, the estimate and sample size for just that year are used.

MODELS

SAIPE Model for Log Number Poor

Let y_{it} and Cen_i denote CPS and census estimates of log(number of poor related children, 5-17). The 1993 SAIPE model (using CPS data for income year 1993) is

(1)

. (2)

The model errors w_{it} and $\tilde{z}_i$ are both assumed *i.i.d.* $N(0, \sigma_w^2)$ and independent of each other.[1] The basic regression variables $\mathbf{x}_{it}$ are defined below. Recall that e_{it} and $\in_i$, the sampling errors in y_{it} and Cen90_i, are assumed independent

[1]Assuming w_{it} independent of $\tilde{z}_i$ is not entirely necessary, but serves as a partial justification for fitting equations (1) and (2) separately. The normality assumption stated here and for other models is also not entirely necessary, as the model fitting and smoothing procedures used can be justified without it.

$N(0, v_e / n_{it})$ and $N(0, c_i)$, with c_i and n_{it} known, and v_e a parameter to be estimated. The unknown parameters to be estimated in (1) and (2) are thus the regression parameters β, γ, η, and $\tilde{\gamma}$; the common model error variance σ_w^2; and the sampling error variance parameter v_e. Decennial census sampling error variances for estimates of number of poor are available from published formulas (generalized variances). If $R_i = \exp(\text{Cen90}_i)$ is the census estimated number of poor, then from a Taylor series linearization, c_i, the sampling error variance in Cen90_i, is approximately

. (3)

Actually, a slight refinement of (3), based on properties of the lognormal distribution was used, as described by Fisher (1997). Practically speaking, the results are not materially different from (3).

The key distinguishing feature of the SAIPE model is the use of the previous census data as a regression variable—the γCen90_i term in (1) and the $\tilde{\gamma}\,\text{Cen80}_i$ term in (2). This SAIPE model form contrasts with the bivariate model form, discussed in the next section. In the SAIPE model form the model error variance, denoted here by σ_w^2, can be essentially thought of as $\text{Var}(Y_i \mid \mathbf{x}_i, \text{Cen90}_i)$, which differs from the model error variance for the bivariate model form, $\sigma_u^2 = \text{Var}(Y_i \mid \mathbf{x}_i)$. The two are not comparable; one would expect $\sigma_w^2 < \sigma_u^2$.

The 1989 SAIPE model (using CPS data for income year 1989) is

(4)

, (5)

with $t = 1989$. Notice that $\mathbf{x}_{it} = \mathbf{x}_{i,89}$, and the regression variables in (4) and (5) are the same. The regression parameters, (β, γ) and $(\eta, \tilde{\gamma})$, are still allowed to be different, however. The same assumptions as above are made about the model errors. Assuming that w_{it} and $\tilde{z}_i$ are independent makes less sense here, since both equations refer to the same year and Cen90_i does not enter (4) as a regression variable. Fortunately, this assumption is unnecessary. Since (4) and (5) contain "identical explanatory variables," regression fitting of these two equations separately produces the same results as fitting them jointly (Theil, 1971:309-310). Finally, notice that the second (census) equations of both the 1993 and 1989 SAIPE models—(2) and (5)—must be the same. Although it might be more appropriate for the 1989 model to replace (5) by the corresponding equation for Cen80_i, this cannot be done because the required regression variables $\mathbf{x}_{i,79}$ are not available.

For this and other models of log(number poor), the CPS estimates y_{it} are defined using 3 years of CPS data for each county i as follows:

$$y_{it} = \log([\text{3-yr weighted avg poverty rate}] \times [\text{3-yr weighted avg poverty universe}]). \qquad (6)$$

The weights given to data from years $t - 1$, t, and $t + 1$ for the weighted averages in (6) are proportional to the numbers of interviewed housing units in county i that contain at least one child aged 5-17 for the year in question. The CPS poverty rate in (6) for county i in year j (= $t - 1$, t, $t + 1$) is

$$\frac{\text{estimated number poor related children 5–17 in county } i\text{, year } j}{\text{estimated total related children 5–17 (CPS poverty universe) in county } i\text{, year } j} \qquad (7)$$

Note that the second term in (6) is the 3-year weighted average of the denominators in (7) for $j = t - 1, t, t + 1$. The CPS poverty universe, and the number of poor related children aged 5-17, are estimated from CPS data for each year using CPS weights modified to make each county "self-representing."

For counties with a CPS sample in only 1 or 2 of the 3 years, the values for only that year, or for the 2-year average corresponding to (6), are used. For counties with no poor children observed in the CPS sample, the direct CPS estimate of the number of poor children is 0. Since logarithms cannot be taken when the direct estimate is 0, y_{it} is not defined, and these counties must be dropped from the model fitting. The same problem arises with the census data, though only for a few counties.

The basic regression variables, $\mathbf{x}_{it} = (x_{0it}, ..., x_{4it})'$, are defined as follows, all but x_{0it} derived from tabulating certain data for each county i:

x_{0it} = 1 (constant term)
x_{1it} = log (number of IRS dependent child tax exemptions on tax returns with income below poverty);
x_{2it} = log (number of food stamp program participants) (from USDA);
x_{3it} = log (resident population aged 0-21);
x_{4it} = log (number of IRS total dependent child tax exemptions). (8)

More recently, Census Bureau analysts have experimented with changing the age limits defining x_{3it} to 0-17. This removed some bias found in evaluations and regression diagnostics for counties with high group quarters populations (usually because of college dorms and military barracks).

Bivariate Model for Log Number Poor

Let y_{it} and Cen_i denote estimates of log(number of poor), as above. The bivariate model form is

(9)

. (10)

The model errors u_{it} and z_i are both *i.i.d.* $N(0, \sigma_u^2)$, with $\text{Cov}(u_{it}, z_i) = \sigma_{uz} = \rho\sigma_u^2$ constant over i. This is the "constrained" bivariate model. The "unconstrained" bivariate model, allowing $\text{Var}(z_i) \equiv \sigma_z^2 \neq \sigma_u^2$, was investigated and found to produce unreasonable results, and it is not considered further here. As above, the sampling errors e_{it} and $\in_i$ are assumed independent $N(0, v_e / n_{it})$ and $N(0, c_i)$, with c_i and n_{it} known, and v_e a parameter to be estimated. Parameters in (9) and (10) to be estimated are thus the regression parameter vectors β and η; the common model error variance σ_u^2; the model error correlation ρ; and the sampling error variance parameter v_e.

Note that the bivariate model form differs from the SAIPE model form in that it does not include the previous census data as a regression variable, and it also allows the model errors to be correlated. These two differences in model form are related. In fact, by making a linear transformation, one could replace (9) by

, (11)

where

(12)

(13)

Replacing (9) by (11) makes the bivariate model form look more like the SAIPE model form, in that both now have the census data on the right-hand side of the CPS equation, and the model errors of the two equations are now uncorrelated. The two differences between (11) and (1) are that (11) uses the regression residuals $\text{Cen}_i - \mathbf{x}'_{i,89}\eta$ instead of just Cen_i, and that γ_i and $\text{Var}(\tilde{w}_{it})$ for (11) vary over counties i. The latter feature makes (11) inconvenient for model estimation

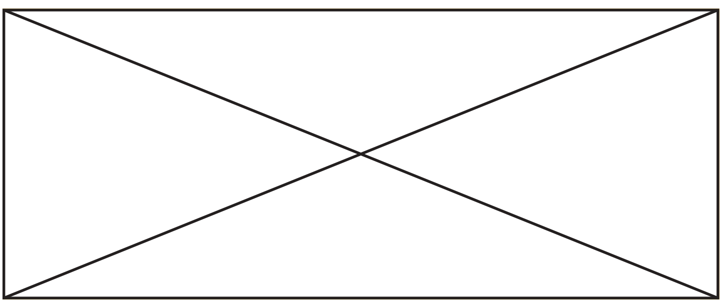
[2]More details related to this transformation of the bivariate model are given in Bell (1997a). To interpret (11), it may help to note that $\mathbf{x}'_{it}\beta + \gamma_i(\text{Cen}_i - \mathbf{x}'_{i,89}\eta) = E(Y_{it}|\mathbf{x}_{it}, \text{Cen}_i)$ and $\text{Var}(\tilde{w}\text{ it}) = \text{Var}(Y_{it}|\mathbf{x}_{it}, \text{Cen}_i)$.

relative to (9). However, having fitted a bivariate model using (9) and (10), one can compute estimates of γ_i and Var($\tilde{w}_{it}$) and compare them to the corresponding quantities γ and σ_w^2 from the SAIPE model (which assumes they are constant over counties). (Histograms of γ_i and Var($\tilde{w}_{it}$) are provided as part of the regression diagnostics for the fitted bivariate models.)[2]

Because the bivariate model uses previous census data Cen_i by jointly modeling it with the CPS data y_{it}, it could not be applied for t = 1989 because the regression variables $\mathbf{x}_{i,79}$ needed for modeling the 1980 census data are not available. Consequently, the bivariate model was applied only for t = 1993, and Cen_i in (10) always denotes Cen90_i. (The bivariate model approach can be applied to jointly model 1990 CPS and 1990 census data, but this is a different exercise, since the resulting smoothed estimates of Y_{it} would use current year census data, rather than previous census data.)

Adding Fixed State Effects to Models

Any of the basic models discussed here can be augmented to include fixed state effects by replacing $x_{0it} = 1$ by a set of 51 state indicator variables, constructed alphabetically: $I_{1i} = 1$ for all counties in Alabama and 0 otherwise, $I_{2i} = 1$ for all counties in Alaska and 0 otherwise, etc., through $I_{51,i} = 1$ for all counties in Wyoming and 0 otherwise. The resulting regression effect can be written as $\sum_{j=1}^{51} \alpha_j I_{ji}$, where the α_j are state intercept parameters. Alternatively, the regression can be reparameterized as follows to maintain the overall constant term $\beta_0 x_{0it}$, but with 50 state contrast variables added to the regression variables for each equation:



where $\boxtimes = \beta_0 = (1/51)\ \sum_{j=1}^{51} \alpha_j$ is the mean of the 51 state intercepts; $\boxtimes_j = \alpha_j - \boxtimes$ are the differential state effects; and $M_{ji} = I_{ji} - I_{51,i}$ are 50 contrast variables that are 1 when county i is in state j, -1 when county i is in Wyoming, and 0 otherwise. The differential state effect for Wyoming is $\boxtimes_{51} = -(\boxtimes_i + \ldots + \boxtimes_{50}$, which is obtained from the constraint $\sum_{j=1}^{51} \tilde{\alpha}_j = 0$.

Two sets of state indicator variables (or state contrast variables) are used—one set for the CPS equation and one set for the census equation. These can be denoted I_{jit} (M_{jit}) and $I_{ji,89}$ ($M_{ji,89}$), which lets the state intercepts α_j (or effects $\boxtimes_j$) be distinct for the CPS and census equations. (The two sets of intercepts could be denoted α_{jt} and $\alpha_{j,\mathrm{cen}}$, or the two sets of contrasts could be denoted $\boxtimes_{jt}$ and $\boxtimes_{j,\mathrm{cen}}$.) Thus, adding state effects to a model adds 100 additional parameters, 50 in each of the two equations: this holds even when modeling CPS data for t = 1989, the same income year as for the census. This approach avoids assuming that state effects are the same for the CPS and census data (though I and my colleagues did do some experimentation with common state effects in the bivariate model).

SAIPE and Bivariate Models for Poverty Rates

All the models that have been investigated are of either the SAIPE or bivariate form, with or without fixed state effects; they are simply applied to different data than discussed above. For modeling poverty rates, Cen_i denotes the census estimated poverty rate for county i (for related children, 5-17). The CPS data y_{it} are defined as an aggregate 3-year "poverty rate," using CPS data for years $t - 1$, t, and $t + 1$:

(14)

where $\sum_t$ indicates the 3-year sum over $t - 1$, t, and $t + 1$. The estimated numbers for the numerator and denominator of (14) are produced by using CPS weights modified to make each county "self-representing." CPS sample sizes n_{it} are defined as before.

Notice that the denominator of (14) is not the CPS poverty universe (poor related children 5-17 in families), as it was for the single-year poverty rates defined in (7); rather, it is the CPS total number of children 5-17. This choice of denominator for the "poverty rate" in (14) is necessary because county population estimates are available for all children 5-17, but not for the 5-17 CPS poverty universe (restricted to related children in families). Population estimates corresponding to the denominator of (14) are needed to convert smoothed poverty rate estimates to estimates of the number of poor children.

For some counties with very small CPS sample sizes there may be no related children aged 5-17 observed in the sample. For these counties, the poverty rates are not defined, and they cannot be used in the model fitting. However, it is not necessary to drop counties just because no *poor* 5-17 children are found in the sample, as it is with the models for log number poor and log poverty rate; the poverty rate models use the most CPS observations for model fitting; 304 counties had CPS sample but no poor age 5-17 in the sample in 1993.

The basic regression variables $\mathbf{x}_{it} = (x_{0it}, \ldots, x_{3it})'$ used in poverty rate models are three other rate variables and an intercept, defined as follows:

$$x_{0it} = 1 \text{ (constant term);} \qquad (15)$$

x_{1it} = (number of IRS dependent child tax exemptions on returns with income below poverty)/(total IRS dependent child tax exemptions);
x_{2it} = (number of food stamp participants) / (resident population, all ages);
x_{3it} = (total IRS dependent child tax exemptions) / (resident pop. age 0-21).

Except for the constant term, the numerators and denominators of these variables derive from tabulations of administrative records data or population estimates for county i. It should be noted that for a significant number of counties (292 in 1993 and 82 in 1989) the IRS dependent child exemption "rate," x_{3it}, exceeds 1: this is partly due to errors in geocoding the IRS tax return data, and partly due to differences between IRS and census residence definitions.

Having thus defined the data and regression variables, either the SAIPE model form given by (1) and (2) or the bivariate model form given by (9) and (10) can be used for the estimates. In doing so, the same assumptions about the error structure are used. Thus, for SAIPE poverty rate models, the model errors w_{it} and $\tilde{z}_i$ in (1) and (2) are both assumed *i.i.d.* $N(0, \sigma_w^2)$ and independent of each other. For bivariate poverty rate models, both model errors u_{it} and z_i in (9) and (10) are assumed *i.i.d.* $N(0, \sigma_u^2)$, with $\text{Cov}(u_{it}, z_i) = \sigma_{uz} = \rho\sigma_u^2$ constant over i. And for both SAIPE and bivariate models the CPS sampling errors e_{it} are assumed *i.i.d.* $N(0, v_e / n_{it})$, and the census sampling errors ϵ_i are assumed *i.i.d.* $N(0, c_i)$. Obviously, the values of the variance parameters will be different from those in the log number poor models: in particular, the census sampling error variances c_i are obtained from published census generalized variances for rate estimates.

To assume that the CPS sampling errors of direct poverty rate estimates have variance of the form v_e / n_{it} is inconsistent with making the same assumption for CPS direct estimates of log number poor or log poverty rate. Simple Taylor series approximations suggest that if v_e / n_{it} is the appropriate variance for poverty rate estimates, then the sampling error variance for log poverty rates will depend on the underlying true poverty rate p, and vice versa. (The sampling error variance for log poverty rates will be the same as that for log number poor, ignoring, as a crude approximation, variability in the denominator of the poverty rates.) In fact, considerations of the binomial distribution suggest that sampling error variances of poverty rates and log poverty rates could both depend on p (see Bell (1997b) for a little more discussion.) The form v_e / n_{it} of the sampling error variances was chosen not because it was believed to be exactly correct for any of the various data being modeled (poverty rates, log poverty rates, or log number poor), but because it is the simplest form that allows sampling error variance to depend inversely on sample size. Because of the need to estimate v_e from the fitting of the CPS equation, it is doubtful that much more involved sampling error variance formulations could be effectively estimated. Since the Census Bureau now has direct estimates of county sampling error variances (Fay, 1997b), there is more information for exploring alternative sampling variance formulations, and that work has begun. (Fixed state effects can also be added to the poverty rate models, as discussed above.)

SAIPE and Bivariate Models for Log Poverty Rates

Models for log poverty rates are of the same form as those for poverty rates just discussed, except that the models are applied with the logarithms of all the rates involved. That is, y_{it} and Cen_i are defined to be the logarithms of the CPS and census poverty rates (defined above) and $(x_{1it}, \ldots, x_{3it})$ are defined to be the logs of the rates given in (15). The y_{it} are not defined for counties for which there are no poor children 5-17 in the CPS sample, so they must be dropped from the model fitting, as is done with the log number poor models.

As with the models discussed above, the assumptions about the covariance structure of (1) and (2) (for a SAIPE model of log poverty rates), or about the covariance structure of (9) and (10) (for a bivariate model), remain unchanged. The parameter values will change, of course: in particular, the sampling variances c_i, which now refer to the log census poverty rates, can be approximated from those for the census poverty rates. Thus, if are the sampling variances in census estimates $\hat{p}_i$ of poverty rates p_i, and c_i are the corresponding sampling variances in the $\text{Cen}_i = \log(\hat{p}_i)$, from Taylor series linearization the two are approximately related by

D-Revised Models for Log Poverty Rates

The "D-Revised" models for log poverty rates are a hybrid: they use CPS and census log poverty rates for y_{it} and Cen_i, as defined above, but with regression variables as defined for the log number poor models in (8).[3] Only the SAIPE form of this model was tried, and fixed state effects were not used. (Alternatives using the bivariate model form or fixed state effects, or both, could be investigated.) For the D-Revised model form there is one additional difference between (1) and (2): the census data appearing on the right-hand side of the equations are—analogous to the other regression variables—defined as log number poor children 5-17, whereas Cen90_i appearing on the left-hand side is the log census poverty rate. With the data thus defined, the model fitting proceeds in the same fashion as for the other models discussed.

State Poverty Rate Models

Models for state poverty rates are discussed in detail in Fay and Train (1997). Here I provide only brief summary remarks relating their model to the forms just discussed. The model developed was of the form of (11), but with the coefficient (γ_i) on the census residuals assumed constant over states i:

[3] "D-Revised" was the term originally used by the panel for the hybrid log rate-number model.

. (16)

The model error variance, $\mathrm{Var}(\tilde{w}_{it}) = \sigma_w^2$, was also assumed constant over states. For states, the census sampling error variances c_i are effectively 0. Thus, examining (12) and (13) for states, a bivariate model does indeed lead to the model form (16), with a constant γ and σ_w^2. In Fay and Train (1997), the equation (16) and corresponding census equation of form (10) were fitted separately. Because the census data have negligible sampling error variance, the census equation for states can be fitted by OLS. Fay and Train then fitted (16) by maximum likelihood to estimate β, γ, and σ_w^2, given previous estimates of the $\mathrm{Var}(e_{it})$.

The estimates of $\mathrm{Var}(e_{it})$ were developed by Mark Otto and myself (see Otto and Bell, 1995). These estimates used generalized variance functions fitted to direct estimates of state sampling error variances developed in Fay and Train (1995). In their later paper on the state modeling, Fay and Train (1997) refined the estimates of $\mathrm{Var}(e_{it})$ as their iterative estimation proceeded by updating the dependence of the $\mathrm{Var}(e_{it})$ on the poverty rate being estimated.

MODEL FITTING

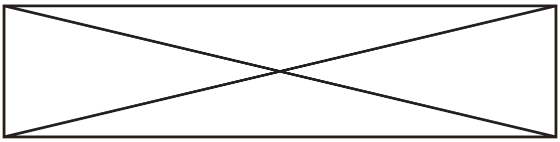
Once the data for a given model have been defined, model fitting proceeds in the same fashion for all models. Thus, model fitting can be discussed in general terms, with one qualification: for models for log number poor or log poverty rates, counties with no CPS sample poor are omitted from the model fitting, as discussed above. Small numbers of other counties may also be eliminated due to no census sample poor or problems in defining the regression variables.

First, consider estimation of the regression parameters given estimates of the model variance parameters. Let **y** and **Cen** (similarly, **Cen90** and **Cen80**) be vectors containing the county CPS and census data to be used for model fitting, and let $\mathbf{X}_t$ and $\mathbf{X}_{89}$ be the corresponding matrices of regression variables for their respective equations. The SAIPE model form given by (1) and (2) can be written in a rather obvious matrix-vector notation as

. (17)

The error vectors $\mathbf{w}_t$, $\tilde{\mathbf{z}}$, $\mathbf{e}_t$, and $\in$ are all assumed uncorrelated with each other, and there are also no correlations among their elements (i.e., each has a diagonal covariance matrix). Thus, Var($\mathbf{w}_t$ + $\mathbf{e}_t$) = $\sigma_w^2 I + v_e K$, where K is a diagonal matrix with elements $1/n_{it}$. Also, Var($\tilde{\mathbf{z}} + \in\ = \sigma_w^2 I + C$, where C is a diagonal matrix with elements c_i. Given σ_w^2, v_e, and the n_{it} and c_i (always assumed known), (17) can be fitted by weighted least squares to estimate the regression parameters ($\beta, \gamma, \eta\, \tilde{\gamma}$). In fact, since there is no correlation between the error terms in the equations for **y** and **Cen90**, these two equations can be fitted separately.

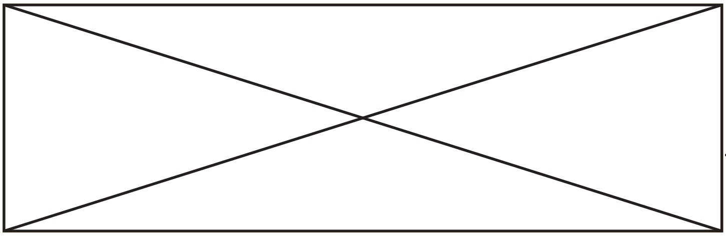
For the bivariate model, the corresponding equation to (17) is



(18)

In (18) the vectors $\mathbf{u}_t$ and $\mathbf{z}$ have, in general, nonzero correlations for observations corresponding to the same county. Thus, while Var($\mathbf{w}_t$ + $\mathbf{e}_t$) = $\sigma_u^2 I + v_e K$ and Var($\mathbf{z}$ + $\in$) $= \sigma_u^2 I + C$, similar to the SAIPE model (17), one also needs to allow for the correlations between $\mathbf{u}_t$ and $\mathbf{z}$ when estimating the regression parameters (β, η). This can be done by applying generalized least squares to (18). In fact, it is simpler to structure the equations for the bivariate model so that the CPS and census data are paired off (for those counties with CPS data available for model fitting), for which the covariance matrix for the resulting equation is block diagonal, with blocks no larger than 2 × 2. (For counties with only census data available for model fitting, the "block" is a scalar.) (This process is straightforward, but the notation is tedious and details are omitted here.)

Fixed state effects are easily added to (17) or (18) by simply augmenting the regression matrix and parameter vector as appropriate. For example, for the bivariate model (18), with 50 state contrast variables M_{ji} and corresponding parameters $_j$ added to each equation, the resulting model can be written

.

Finally, it is necessary to discuss how the covariance parameters are estimated and how this estimation is integrated with that for the regression parameters. Two approaches have been taken. One approach (implemented in SAS IML) was used in fitting models to produce the evaluations against the 1990 census. This approach used basically a method of moments approach (see Fisher, 1997).

The second approach (implemented in Splus) was used in fitting the models for producing the regression diagnostics. This approach uses Gaussian maximum likelihood. For bivariate form models, for given values of the model parameters (β, η, σ_u^2, ρ, v_e), the joint density of the data (the likelihood function) can be evaluated, and thus numerically maximized over the parameters to produce the maximum likelihood estimates (MLEs). This is done by iterating between GLS estimation of (β, η) for given values of (σ_u^2, ρ, v_e) and maximization of the likelihood over (σ_u^2, ρ, v_e), using the regression residuals $y_{it} - \mathbf{x}'_{it}\beta$ and $\text{Cen}_i - \mathbf{x}'_{i\,89}\eta$ as data. This approach can be called iterative GLS. Asymptotic inference (approximate standard errors, etc.) about (β, η) follows from standard GLS results by plugging in MLEs of (σ_u^2, ρ, v_e), and inference about (σ_u^2, ρ, v_e) uses standard asymptotic results for MLEs (use of an approximate normal distribution with covariance matrix given by the inverse negative Hessian of the log-likelihood evaluated at the MLEs).

This second approach can also fit models of the SAIPE form. For these models, $\rho = 0$, so the CPS and census equations are independent. However, these two equations are linked by the common variance, σ_w^2, assumed for the model errors w_{it} and $\tilde{z}_i$. Thus, fitting the two equations jointly combines their information for the estimation of σ_w^2. Practically speaking, this makes little difference, as the information from the census data swamps that from the CPS data, so that essentially the same results would be obtained by fitting the census equation first to estimate σ_w^2 and then treating σ_w^2 as known when estimating the CPS equation. This latter strategy was used in the first approach (implemented in SAS IML).

The SAS program differs from the Splus program in another related respect: in the SAS program the census equation is fitted only to data from the counties that also provide data for the CPS equation. The reasoning behind this decision was that the model error variance might differ for counties without a CPS sample (which are smaller, on average, than counties included in the CPS), and thus it may be appropriate to exclude them from the fitting of the census equation. As noted in the next section, an important role of the model error variance relates to how weights are assigned to the regression predictions and the direct CPS estimates in constructing the smoothed estimates. Since this calculation is irrelevant to counties without a CPS sample, it may be appropriate to avoid their influence on estimates of the model error variance. In the Splus bivariate model software, all the census data are used in the model fitting, along with as much CPS data as are available for the year and the poverty statistic being modeled. This approach assumes that the model applies equally well to counties with and without a CPS sample.

The two different model fitting approaches were adopted because some analysts use SAS and others use Splus and because the SAS code was developed for the original SAIPE model and could not be used to fit models of bivariate form, necessitating development of a second program. Generalization of the Splus bivariate model software is a recent development, and there has not been time to make extensive comparisons of the two programs for models they can both fit. For the comparisons that have been made, the differences in results appear to be small.

SMOOTHED ESTIMATES

Smoothed estimates from an estimated 1993 SAIPE model form are determined from the CPS equation (1), treating Cen90_i the same way as the other regression variables in $\mathbf{x}_{it}$. (For t = 1989, the same approach is applied to (4).) Recall that the true quantity of interest for county i is $Y_{it} = \mathbf{x}'_{it}\beta + \gamma\text{Cen90}_i + w_{it}$, and the direct CPS estimate is $y_{it} = Y_{it} + e_{it}$. The estimate of Y_{it} and its variance are

(19)

, (20)

where

$$h_{it} = \sigma_w^2/(\sigma_w^2 + v_e/n_{it}),$$

and Var (,) is obtained from the weighted least squares results. From (19) the smoothed estimate $\hat{Y}_{it}$ is a weighted average of the regression prediction $\mathbf{x}'_{it}$ + Cen90_i and the direct estimate y_{it}. The first term in (20), $\sigma_w^2(1 - h_{it})$, is the variance that would result if all model parameters were known. The second term in (20) accounts for additional error due to estimating the regression parameters (β, γ). One can also augment (20) to account for additional error due to estimating some or all of the variance parameters (σ_w^2 and v_e), using either the approach of Prasad and Rao (1990:47-59), or by simulation. These calculations have been done for some of the models, and this addition to the variance was found to be small. (Note that the models have a small number of variance parameters relative to the amount of data.)

For models with fixed state effects, smoothed estimates and their variances are obtained from expressions analogous to (19) and (20) by appropriately augmenting the regression variables and parameters with the state effect regression variables and parameters.

For counties without a CPS sample or that have a CPS sample with no poor children and are dropped from the fitting of log(number poor) or log(poverty rate) models, the estimate $\hat{Y}_{it}$ is defined to be just the regression prediction $\mathbf{x}'_{it}\hat{\beta} + \hat{\gamma}\,\text{Cen90}_i$, which has variance

$$\text{Var}(Y_{it} - \hat{Y}_{it}) = \sigma_w^2 + [\mathbf{x}'_{it}\ \text{Cen90}_i]\ \text{Var}\begin{pmatrix}\hat{\beta} \\ \hat{\gamma}\end{pmatrix}\begin{bmatrix}\mathbf{x}_{it} \\ \text{Cen90}_i\end{bmatrix}.$$

Smoothed estimates and their variances for the bivariate model are a little more complicated, but follow the same principles; they are discussed in Bell (1997a).

When log(numbers of poor) or log(poverty rates) are modeled, smoothed estimates on the original scale (of numbers of poor or of poverty rates, unlogged) can be obtained by exponentiating $\hat{Y}_{it}$. However, it is useful to use the following modified estimate, based on the mean of the lognormal distribution, to remove bias:

$$\exp\left(\hat{Y}_{it} + \frac{1}{2}\text{Var}(Y_{it} - \hat{Y}_{it})\right). \tag{21}$$

Prediction intervals on the original scale can be obtained by exponentiating prediction interval limits on the transformed (log) scale, yielding asymmetric intervals on the original scale.

When poverty rates are modeled, the resulting smoothed rate estimate for county i must be multiplied by the population estimate of total children 5-17 in county i (see (14) and discussion following) to convert it to a smoothed estimate of the number of poor children. This is also necessary for smoothed poverty rate estimates from the state model, and, similarly, when log(poverty rates) for counties are modeled, with smoothed rate estimates produced using (21). Prediction error variances in these cases could be taken to be those for the smoothed poverty rates multiplied by the square of the population estimates, though this ignores error in the 5-17 population estimates. Formal measures (variances) of error in state and county population estimates are not available, so there is no ready way to recognize this additional uncertainty. Treating error in the population estimates as ignorable is more tenable for states than it is for counties.

As a final step, smoothed county estimates of number of poor related children aged 5-17 are "raked" to agree with the corresponding smoothed estimates from the state model. Thus, the smoothed county estimates are aggregated to states, and then the individual county estimates are multiplied by the ratio of their state model estimate to the aggregated county estimates for that state. These ratios, or "raking factors," one for each state for a given model, have been developed for the 1989 models. Deriving variances for the raked, smoothed estimates is complicated, but an approximate procedure (described in Fisher, 1997) has been implemented in conjunction with the SAS estimation software.

APPENDIX

B

Regression Diagnostics on Alternative County Regression Models

An internal evaluation of a regression model, or "regression diagnostics," involves an assessment of its underlying assumptions and features. Chapter 6 reports the results of such an evaluation for four county models, estimated for 2 years, 1989 and 1993. These four models, which were considered serious candidates to produce revised county estimates of poor school-age children in 1993, have the following designations: (a) log number model (under 21, the original county model); (b) log number model (under 18, the revised county model); (c) log rate model (under 21); and (d) log rate model (under 18).

This appendix summarizes the results of an internal evaluation for 13 county models, listed below (see Chapter 5 and Appendix A for the model specifications). Twelve of the models were considered in the first round of model evaluations; they include models (a), (b), and (c). The other model, the log rate (under 18) model (d), was added for the second round of evaluations, which considered the four candidate models (a-d).

Of the 13 county models, 7 are single-equation models, in which the dependent variable is from 3 years of the CPS. For 1993 estimates of poor school-age children, the dependent variable is a weighted average of data from the March 1993, 1994, and 1995 CPS, covering income years 1992, 1993, and 1994. For 1989 estimates of poor school-age children, produced for evaluation purposes, the dependent variable is a weighted average of data from the March 1989, 1990, and 1991 CPS, covering income years 1988, 1989, and 1990.

The other 6 county models are bivariate models in which two equations are jointly estimated to develop estimates of poor school-age children in 1993. In one equation, the dependent variable is a weighted average of data from the March 1993, 1994, and 1995 CPS, covering income years 1992, 1993, and 1994. In the second equation, the dependent variable is from the 1990 census, covering income year 1989.

The regression coefficients for all the CPS models are presented in Table B-1; Table B-2 shows the regression coefficients for the 1990 census equation for the 6 bivariate models (see pages 188-190).

Single-Equation Models	**Bivariate Models**
Log number under 21 (1989, 1993)	Log number under 21 (1993)
Log number under 18 (1989, 1993)	
Log number under 21, fixed state effects (1989, 1993)	Log number under 21, fixed state effects (1993)
Log rate under 21 (1989, 1993)	Log rate under 21 (1993)
	Log rate under 21, fixed state effects (1993)
Log rate under 18 (1989, 1993)	
Rate under 21 (1989, 1993)	Rate under 21 (1993)
	Rate under 21, fixed state effects (1993)
Hybrid log rate-number under 21 (1989, 1993)	

NOTE: The years for which coefficients were fit are in parentheses; for the bivariate models, the year shown is for the CPS equation.

REGRESSION DIAGNOSTICS METHODS

Regression diagnostics is an analysis of the extent to which the various assumptions on which a regression model is based are supported by the data. The following six assumptions were examined for the 13 county models of poor school-age children (see Chapter 6):

(1) linearity of the relationship between the dependent variable and the predictor variables;
(2) constancy over time of the assumed linear relationship and in the estimated coefficients of the predictor variables;
(3) which variables are needed in the model, specifically, whether any of the included predictor variables are *not* needed in the model and, con-

versely, whether other potential predictor variables *are* needed in the model;

(4) normality (primarily symmetry and moderate tail length) of the distribution of the standardized residuals;[1]

(5) whether the standardized residuals have homogeneous variances; that is, whether the variability of the standardized residuals is constant across counties and does not depend on the values of the predictor variables; and

(6) the absence of outliers, which can be considered to be the absence of an extremely long tail to the error distribution.

Various techniques are useful for examining the degree to which each of these six assumptions obtain. The following techniques that were implemented by the panel and the Census Bureau to evaluate the 13 county models are certainly not the only ones that can be used to examine each of the above assumptions, but they are usually included. In addition to these general techniques, specific analyses were conducted to evaluate the bivariate model formulation in comparison with the single-equation model formulation and the use of the population under age 18 in comparison with the population under age 21 as a predictor variable in the log number model.

Linearity Linearity of the relationships between the dependent variable and the predictor variables was assessed graphically, by observing whether there was evidence of curvature in the plots of standardized residuals against predictor variables in the model. In addition, plots of residuals against CPS sample size and against the predicted values from the regression model were examined for curvature.

Constancy For the single-equation models that could be fit for both 1989 and 1993, the regression coefficients were compared to determine if the values remained roughly constant over time.

Inclusion or Exclusion of Predictor Variables The possibility that one or more predictor variables should be excluded from a model was assessed by looking for insignificant *t*-statistics for the estimated values of individual regression coefficients. The need to include additional predictor variables was assessed by looking for nonrandom patterns, indicative of possible model bias, in the distributions of standardized residuals displayed for various categories of counties. (See Chapter 6 for the categories examined in various model evaluations;

[1]See Chapter 6 for the procedure used to standardize the residuals, which are the differences between the predicted and reported values of the dependent variable for each observation.

TABLE B-1 Estimates of Regression Coefficients for the CPS Equation for 13 County Models

	Predictor Variables[a]				
Model	1	2	3	4	5
Log Number (under 21)					
1989	0.52	0.30	0.76	–0.81	0.27
	(.07)	(.05)	(.22)	(.22)	(.07)
1993	0.31	0.30	0.03	0.03	0.40
	(.08)	(.07)	(.21)	(.21)	(.09)
Log Number (under 18)					
1989	0.50	0.23	1.79	–1.80	0.32
	(.06)	(.05)	(.27)	(.27)	(.07)
1993	0.38	0.27	0.65	–0.59	0.34
	(.08)	(.07)	(.24)	(.24)	(.09)
Log Number (under 21), Fixed State Effects					
1989	0.36	0.27	0.45	–0.56	0.51
	(.13)	(.07)	(.25)	(.25)	(.10)
1993	0.50	0.17	–0.03	–0.07	0.45
	(.12)	(.09)	(.25)	(.25)	(.11)
Hybrid Log Rate-Number (under 21)					
1989	0.55	0.27	0.35	–1.34	0.25
	(.06)	(.05)	(.21)	(.21)	(.06)
1989	0.37	0.26	–0.33	–0.59	0.37
	(.07)	(.06)	(.18)	(.18)	(.08)
Bivariate Log Number (under 21)					
1993	0.57	0.45	0.19	–0.20	NA
	(.06)	(.05)	(.20)	(.20)	
Bivariate Log Number (under 21), Fixed State Effects					
1993	0.83	0.34	0.21	–0.38	NA
	(.09)	(.07)	(.24)	(.24)	

TABLE B-1 Continued

	Predictor Variables[b]			
Model	1	2	3	4
Log Rate (under 21)				
1989	0.32	0.29	–0.73	0.40
	(.07)	(.04)	(.19)	(.07)
1993	0.23	0.31	–0.07	0.41
	(.08)	(.06)	(.18)	(.09)
Log Rate (under 18)				
1989	0.29	0.26	–1.13	0.43
	(.07)	(.04)	(.24)	(.07)
1993	0.26	0.30	–0.42	0.38
	(.08)	(.06)	(.20)	(.09)
Rate (under 21)				
1989	0.25	0.46	–0.16	0.56
	(.06)	(.08)	(.03)	(.06)
1993	0.09	0.60	–0.05	0.52
	(.06)	(.11)	(.03)	(.10)
Bivariate Log Rate (under 21)				
1993	0.57	0.40	–0.12	NA
	(.05)	(.04)	(.16)	
Bivariate Log Rate (under 21), Fixed State Effects				
1993	0.75	0.35	–0.01	NA
	(.08)	(.05)	(.19)	
Bivariate Rate (under 21)				
1993	0.38	0.89	–0.05	NA
	(.04)	(.06)	(.03)	
Bivariate Rate (under 21), Fixed State Effects				
1993	0.44	0.85	–0.05	NA
	(.06)	(.08)	(.04)	

NOTES: All predictor variables are on the logarithmic scale for numbers and rates. Standard errors of the estimated regression coefficients are in parentheses. Estimated coefficients for the state indicator variables are not shown. The models were estimated with maximum likelihood. NA: not applicable.

[a]Predictor variables: (1) number of child exemptions reported by families in poverty on tax returns (1989 or 1993); (2) number of people receiving food stamps (1989 or 1993); (3) population (under age 21 or under age 18, 1990 or 1994); (4) total number of child exemptions on tax returns (1989 or 1993); (5) number of poor school-age children from previous (1980 or 1990) census.

[b]Predictor variables: (1) ratio of child exemptions reported by families in poverty on tax returns to total child exemptions (1989 or 1993); (2) ratio of people receiving food stamps (1989 or 1993) to total population; (3) ratio of total child exemptions on tax returns (1989 or 1993) to population (under age 21 or under age 18); (4) ratio of poor school-age children from previous (1980 or 1990) census.

TABLE B-2 Estimates of Regression Coefficients for the 1990 Census Equation for the 1993 Bivariate Models

Model	Predictor Variables[a]			
	1	2	3	4
Bivariate Log Number (under 21)	0.71	0.31	0.48	–0.51
	(.01)	(.01)	(.03)	(.03)
Bivariate Log Number (under 21), Fixed State Effects	0.71	0.33	0.45	–0.48
	(.02)	(.01)	(.03)	(.03)
	Predictor Variables[b]			
Bivariate Log Rate (under 21)	0.66	0.30	–0.23	N.A.
	(.01)	(.01)	(.02)	
Bivariate Log Rate (under 21), Fixed State Effects	0.67	0.30	–0.22	N.A.
	(.01)	(.01)	(.02)	
Bivariate Rate (under 21)	0.56	0.75	–0.05	N.A.
	(.01)	(.01)	(.01)	
Bivariate Rate (under 21), Fixed State Effects	0.55	0.78	–0.05	N.A.
	(.01)	(.02)	(.01)	

NOTE: See notes to Table B-1.

[a]Predictor variables: (1) number of child exemptions reported by families in poverty on tax returns in 1989; (2) number of people receiving food stamps in 1989; (3) population under age 21 in 1990; (4) total number of child exemptions on tax returns in 1989.

[b]Predictor variables: (1) ratio of child exemptions reported by families in poverty on tax returns to total child exemptions in 1989; (2) ratio of people receiving food stamps in 1989 to total population; (3) ratio of total child exemptions on tax returns in 1989 to population under age 21.

the distributional displays examined for this and other model assumptions were box plots.)

Normality The normality of the standardized residuals was evaluated through use of Q-Q plots, histograms, and box plots of the standardized residuals. While some skewness of the distribution of standardized residuals may be acceptable, extreme skewness can change the regression fit so that a relatively small number of counties have more influence on the estimation of the regression coefficients. In addition, extreme skewness can indicate the need for a transformation of the variables, which might, in turn, reveal the need for additional predictor variables.

Homogeneous Variances The homogeneity of the variance of the standardized residuals was assessed using several statistics and graphical displays. The statistics included: Spearman's rank correlation coefficient of absolute standardized residuals with the predicted values and also with the CPS sample size, and a robust regression of the log absolute standardized residuals on CPS sample size. The graphical displays included: scatterplots of absolute standardized residuals versus model predictor variables; box plots of absolute standardized residuals for categories of counties; plots of the median absolute deviation of the standardized residuals in a category by categories; plots of absolute standardized residuals versus log CPS sample size; and plots of standardized residuals to the two-thirds power (the Wilson-Hilferty transformation) versus log CPS sample size.

Outliers The assumption of the absence of outliers was evaluated through examination of plots of the distributions of the standardized residuals and plots of standardized residuals against the predictor variables and through analysis of patterns in the distribution of the 30 largest absolute standardized residuals for the various characteristics used to categorize the counties.[2] Any patterns observed among the 30 largest absolute standardized residuals for a characteristic may suggest that a predictor variable should be added to a model.

FINDINGS

Linearity There is no evidence of any strong nonlinearity between the predictor variables and the dependent variable in any of the 13 models. Thus, there is no reason to suggest a transformation of the dependent variable in any of the models, nor is there reason to include any higher order polynomial terms as additional predictor variables.

Constancy The regression coefficients for the 7 single-equation models for 1989 and 1993 are shown in Table B-1. All of these models have some coefficients that differ substantially between 1989 and 1993.

Inclusion or Exclusion of Predictor Variables All of the models with fixed state effects have a large fraction of state effects that are not significant at the 5 percent level. In addition, several other models, especially for 1993, had one or two predictor variables with regression coefficients that were not significant, but that was typically for only 1 of the 2 years that were analyzed. Therefore, except for the models with fixed state effects, there was little evidence of predictor variables that should be excluded from an equation. For the fixed state effects models, an examination of the extent to which the state effects cluster and could

[2]All the outlier statistics examined are based on the residuals from a least squares model fit, so they may miss influential outliers. It would be useful to look for outliers from a robust fit of the models. It would also be useful to compare the predictions from models with extreme outliers removed.

be estimated in groups might make it possible to reduce the number of coefficients that need to be estimated.

With respect to the need to include additional predictor variables in a model, nonrandom patterns of the distributions of the standardized residuals—especially a difference in the median standardized residual from 0 for the residuals in a county category—were observed for several characteristics: percent Hispanic population, location in a metropolitan area outside the central county, and population size. The models with the fewest nonrandom patterns of the distributions of the standardized residuals were the bivariate log rate, bivariate rate, and rate models.

Normality Many of the models had distributions of the standardized residuals that were both asymmetric and long-tailed, especially to the side to which the distribution was skewed. It was difficult to distinguish between skewness and the presence of outliers. Often, movement from a log number dependent variable to a log rate dependent variable reduced an outlier problem, but it introduced a skewness problem. The rate models and the hybrid log rate-number model seemed to have both problems and to be particularly problematic in this respect. In contrast, the log number models behaved relatively well on this criterion.

Homogeneous Variances All of the models exhibited nonconstant variances of the standardized residuals. One would expect the standardized residual variance to remain constant over the distribution of CPS sample size; however, for these models, it increased with increasing sample size. Most of the models also had some variance heterogeneity as a function of the predicted value (number or proportion of poor school-age children).

Outliers The rate models and the hybrid log rate-number model exhibited both skewness and long-tailed error distributions. For all models, large urban counties, particularly those with large percentages of Hispanics, and counties that are in metropolitan areas but not the central county had somewhat more outliers than other counties. The bivariate log rate, bivariate log number, and the log rate models had fewer outliers that demonstrated these patterns.

Additional Analysis Analysis that focused on a regression coefficient that is assumed to be constant in the single-equation formulation and is variable in the bivariate formulation demonstrated strong heterogeneity, thereby supporting the bivariate approach (see Appendix A). Also, Akaike's information criterion (AIC) confirmed the superiority of using the population under age 18 as a predictor variable in the log number model instead of the population under age 21.

SUMMARY

Analysis of the regression output for the 13 county models for the most part supports the assumptions of the models; it does not strongly support one model over another. All of the models exhibit a few common problems. First, they all behave somewhat differently for larger urban counties, especially those with large percentages of Hispanics, than for rural counties. Second, all models show evidence of some variance heterogeneity, particularly with respect to CPS sample size and often with respect to the predicted value (number or proportion of poor school-age children). The rate models and the hybrid log rate-number model exhibit more problems with skewness and outliers than other model formulations. The bivariate approach appears promising due to the heterogeneity in the regression coefficient mentioned above, the lack of patterns in the analysis of the standardized residuals, and the correlation observed by corresponding residuals in the CPS and census regression equations. Finally, according to the internal evaluation, none of the alternative models is clearly superior to the log number model, and the use of the predictor variable for the population under age 18 instead of under age 21 is supported for the log number model.

APPENDIX C
County Model Comparisons with 1990 Census Estimates

An external evaluation of alternative models for producing county estimates of poor school-age children can be carried out by comparing the county estimates obtained from each model for 1989 with 1900 census estimates of related children 5-17 who were poor in 1989. Although this evaluation is not ideal, it serves as a valuable tool for model assessment.

Chapter 6 reports the results of such an evaluation for four candidate models and four procedures that rely more heavily on estimates from the 1980 census. This appendix supplements the material in Chapter 6 in two ways. First, it provides additional results for the four models and four procedures examined in Chapter 6. Second, it provides evaluation results for the six single-equation models that were considered in the first round of evaluations.

EVALUATION MEASURES

Four measures are used for the evaluations in Chapter 6 and in this appendix. Two are overall measures of the differences between the county estimates from a model (or procedure) and the census, and two are measures for categories of counties. The four measures are defined as follows:

(1) *Average absolute difference*: the sum over all counties of the absolute (unsigned) difference between the model estimate of poor school-age children and the 1990 census estimate for each county, divided by the number of counties (3,141), or

$$\Sigma(|Y_{\text{model } i} - Y_{\text{census } i}|) / n .$$

(2) *Average proportional absolute difference*: the sum over all counties of the absolute difference between the model estimate of poor school-age children and the 1990 census estimate as a proportion of the census estimate for each county, divided by the number of counties,[1] or

$$\Sigma[(|Y_{\text{model } i} - Y_{\text{census } i}|) / Y_{\text{census } i}] / n.$$

(3) *Category algebraic difference*: the sum for all counties (i) in a category (j) of the algebraic (signed) difference between the model estimate of poor school-age children and the 1990 census estimate for each county in the category, divided by the sum of the census estimates for the counties in the category, or

$$\Sigma_i \ (Y_{\text{model } ij} - Y_{\text{census } ij}) \ / \Sigma_i Y_{\text{census } ij} .$$

(4) *Category average proportional algebraic difference*: the sum for all counties (i) in a category (j) of the algebraic difference between the model estimate of poor school-age children and the 1990 census estimate as a proportion of the census estimate for each county in the category, divided by the number of counties in the category, or

$$\Sigma_i \ [(Y_{\text{model } ij} - Y_{\text{census } ij}) / Y_{\text{census } ij}] / n_j .$$

Measure (1) expresses overall absolute model-census differences in terms of numbers of poor school-age children; measure (2) expresses overall absolute model-census differences in terms of percentage errors for counties. Similarly, for categories of counties, measure (3) expresses model-census differences in terms of numbers of poor school-age children, while measure (4) expresses model-census differences in terms of percentage errors for counties. The two kinds of category differences are algebraic (not absolute) measures, in which positive differences offset negative differences.

For measures (3) and (4), the counties are grouped into categories of the following characteristics: census geographic division; metropolitan status of county; population size in 1990; population growth from 1980 to 1990; percentage of poor school-age children in the 1980 census; percentage of Hispanic population in 1990; percentage of black population in 1990; persistent poverty from 1960 to 1990 for rural counties; economic type for rural counties; percentage of group quarters residents in 1990; whether the county had households in the CPS sample; and percentage change from 1980 to 1990 in the proportion of poor school-age children.[2] Tables C-1 and C-2 show the number of counties in each category.

[1] An analogous measure, shown in Table 6-3, is the average proportional absolute difference in estimated proportions of poor school-age children.

[2] The characteristic of percentage change in the proportion of poor school-age children from 1980 to 1990 was not included in the first round of evaluations.

TABLE C-1 Comparison of Model Estimates and Other Procedures with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Algebraic Difference by Category of County (in percent)

Category	Counties[a] (Number)	Model: Log No. Under 21 (a)	Model: Log No. Under 18 (b)	Model: Log Rate Under 21 (c)	Model: Log Rate Under 18 (d)
Census Division[b]					
New England	67	–2.9	–2.9	–2.9	–2.9
Middle Atlantic	150	–2.8	–2.8	–2.8	–2.8
East North Central	437	–0.2	–0.2	–0.2	–0.2
West North Central	618	1.7	1.7	1.7	1.7
South Atlantic	591	0.5	0.5	0.5	0.5
East South Central	364	–4.5	–4.5	–4.5	–4.5
West South Central	470	–2.7	–2.7	–2.7	–2.7
Mountain	281	4.3	4.3	4.3	4.3
Pacific	163	6.5	6.5	6.5	6.5
Metropolitan Status					
Central county of metropolitan area	493	2.4	1.6	–0.1	–0.5
Other metropolitan	254	–6.6	–5.0	5.1	6.3
Nonmetropolitan	2,394	–4.2	–2.8	–0.3	0.4
1990 Population Size					
under 7,500	525	–9.0	–2.3	–1.9	2.3
7,500-14,999	630	–4.4	0.5	2.5	5.5
15,000-24,999	524	–5.1	–2.6	0.3	1.9
25,000-49,999	620	–4.2	–2.9	0.6	1.3
50,000-99,999	384	–3.5	–5.1	–1.2	–2.3
100,000-249,999	259	–1.8	–4.4	–1.8	–3.5
250,000 or more	199	3.3	3.2	0.5	0.5
1980 to 1990 Population Growth					
Decrease of more than 10.0%	444	–1.9	0.6	–3.4	–1.9
Decrease 0.1-10.0%	972	–0.6	–0.5	–1.9	–1.8
0.0-4.9%	547	–2.8	–2.8	–3.2	–3.1
5.0-14.9%	620	0.0	–1.0	0.2	–0.6
15.0-24.9%	260	7.7	5.8	5.5	4.6
25.0% or more	292	–4.0	–1.4	1.7	3.1

Other Procedures			
Stable Shares (i)	Stable Shares in State (ii)	Stable Rates in State (iii)	Average of Census and (a) (iv)
35.9	–2.9	–2.9	7.8
27.1	–2.8	–2.8	4.4
–2.8	–0.2	–0.2	–5.6
–1.8	1.7	1.7	–2.1
14.8	0.5	0.5	8.1
14.1	–4.5	–4.5	2.1
–18.1	–2.7	–2.7	–6.3
–23.2	4.3	4.3	–3.1
–21.3	6.5	6.5	0.2
–1.6	–0.6	–0.4	0.4
3.2	–1.6	10.1	3.4
3.3	1.8	–0.5	–1.4
16.5	23.0	9.4	1.3
10.9	10.7	4.4	2.2
6.2	3.4	0.0	–0.6
2.4	–0.2	–0.3	–1.3
–2.5	–4.8	–2.5	–3.3
–4.9	–5.9	–2.9	–3.3
–0.6	0.8	0.8	1.8
9.1	9.9	–3.1	–3.4
7.5	0.7	–4.6	–2.7
11.0	–2.3	–3.3	–0.2
6.1	0.2	1.7	2.1
–12.8	4.4	3.5	2.4
–21.2	–6.8	7.2	1.0

continued on next page

TABLE C-1 Continued

Category	Counties[a] (Number)	Model: Log No. Under 21 (a)	Log No. Under 18 (b)	Log Rate Under 21 (c)	Log Rate Under 18 (d)
Percent Poor School-Age Children, 1980					
Less than 9.4%	516	–4.0	–4.5	0.0	0.2
9.4-11.6%	524	–0.5	–1.0	–1.6	–1.8
11.7-14.1%	530	3.6	2.3	1.8	1.0
14.2-17.2%	523	0.9	1.2	–1.2	–1.4
17.3-22.3%	519	1.8	1.7	0.3	–0.1
22.4-53.0%	523	–2.2	0.8	1.3	2.8
Percent Hispanic, 1990					
0.0-0.9%	1,770	–3.4	–3.3	–1.6	–1.5
1.0-4.9%	847	0.5	0.1	0.4	0.1
5.0-9.9%	193	–1.4	–0.6	–1.1	–0.8
10.0-24.9%	181	2.2	1.8	0.7	0.5
25.0-98.0%	150	3.9	4.6	2.2	2.7
Percent Black, 1990					
0.0-0.9%	1,446	–1.2	0.3	3.9	4.9
1.0-4.9%	615	–0.7	–2.0	1.3	0.5
5.0-9.9%	294	–2.9	–2.5	–0.7	–0.6
10.0-24.9%	381	2.0	1.2	–1.0	–1.3
25.0-87.0%	405	1.0	1.7	–1.8	–1.4
Persistent Rural Poverty, 1960-1990[c]					
Rural, not poor	1,740	–4.0	–3.7	–1.2	–1.0
Rural, poor	535	–5.0	–2.1	0.7	2.1
Not classified	866	1.7	1.2	0.3	0.0
Economic Type, Rural Counties[c]					
Farming	556	–5.5	–2.5	–1.6	0.7
Mining	146	–10.7	–5.1	–6.3	–3.6
Manufacturing	506	–6.2	–5.9	–1.7	–1.0
Government	243	2.1	–1.3	6.3	3.2
Services	323	–3.9	–3.0	–1.8	–1.2
Nonspecialized	484	–3.7	–1.0	–0.1	1.4
Not classified	883	1.7	1.2	0.3	0.0

Other Procedures			
Stable Shares (i)	Stable Shares in State (ii)	Stable Rates in State (iii)	Average of Census and (a) (iv)
2.4	0.8	5.1	–1.1
–9.9	–4.0	–1.9	–3.6
–4.2	1.8	0.7	0.2
–5.0	–3.0	–5.3	–1.8
10.7	1.9	–0.1	4.2
12.3	4.1	1.8	4.1
10.7	–0.6	–1.4	0.2
0.2	0.1	1.1	–0.4
6.7	1.2	1.4	1.7
–5.7	1.7	1.3	0.1
–16.8	–1.2	–1.3	–0.4
–3.7	3.9	6.0	–0.5
–6.3	–1.6	–0.4	–2.9
–8.4	–2.3	2.2	–1.8
–2.6	–0.7	–2.1	0.2
16.5	1.2	–2.4	3.7
0.1	0.2	–1.4	–3.4
9.8	5.4	0.1	1.2
–1.2	–0.7	0.4	0.7
13.2	18.0	7.9	1.1
–8.9	–6.6	–13.1	–10.6
12.1	0.8	–1.1	–0.2
–0.9	4.6	4.1	0.0
–5.8	–4.0	–3.4	–4.3
2.2	1.6	–2.0	–1.5
–1.2	–0.7	0.4	0.7

continued on next page

TABLE C-1 Continued

Category	Counties[a] (Number)	Model: Log No. Under 21 (a)	Log No. Under 18 (b)	Log Rate Under 21 (c)	Log Rate Under 18 (d)
Percent Group Quarters Residents, 1990					
Less than 1.0%	545	–6.7	–2.7	2.0	4.7
1.0-4.9%	2,187	0.3	0.7	–0.3	0.1
5.0-9.9%	299	2.3	–4.4	0.5	–5.2
10.0-41.0%	110	14.2	–3.2	7.4	–7.5
Status in CPS, 1989-1991					
In CPS sample	1,028	1.4	1.0	–0.2	–0.5
In CPS, no poor children 5-17	246	–2.6	–1.9	7.3	7.8
Not in CPS sample	1,867	–4.1	–2.8	–0.1	0.6
Change in Poverty Rate for School-Age Children, 1980-1990					
Decrease of more than 3.0%	536	7.5	10.4	16.2	18.1
Decrease 0.1-3.0%	649	2.1	1.9	3.1	2.9
0.0-0.9%	272	–2.6	–0.8	–0.4	0.5
1.0-3.4%	621	3.8	2.2	3.4	2.6
3.5-6.4%	532	–1.2	–2.4	–3.8	–4.3
6.5-38.0%	523	–7.2	–5.2	–8.7	–7.8

NOTES: The census estimates are controlled to the CPS national estimate for 1989. The algebraic difference by category is the sum for all counties in a category of the algebraic (signed) difference between the model estimate of poor school-age children and the 1990 census estimate for each county, divided by the sum of the census estimates for all counties in the category. See Chapter 6 text for definitions of models.

[a]3,141 counties are assigned to a category for most characteristics; 3,135 counties are assigned to a category for 1980-1990 population growth and 1980 percentage poor school-age children; 3,133 counties are assigned to a category for 1980-1990 percentage change in poverty rate for school-age children.

Other Procedures			
Stable Shares (i)	Stable Shares in State (ii)	Stable Rates in State (iii)	Average of Census and (a) (iv)
–1.4	–0.9	3.7	0.3
–0.4	0.3	–0.1	0.1
7.8	–1.4	–2.8	–0.8
1.8	–0.9	–1.4	–2.2
–0.6	–0.7	–0.4	0.5
10.0	3.7	12.0	5.9
0.6	2.3	–0.3	–2.3
51.6	30.1	32.8	30.0
29.2	8.0	9.8	12.1
4.3	–0.9	3.3	3.1
–5.1	3.7	3.4	0.2
–14.3	–7.7	–9.5	–8.3
–25.2	–14.2	–16.5	–14.5

[b]Census division states:
New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
Middle Atlantic: New York, New Jersey, Pennsylvania
East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin
West North Central: Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, Kansas
South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
East South Central: Kentucky, Tennessee, Alabama, Mississippi
West South Central: Arkansas, Louisiana, Oklahoma, Texas
Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada
Pacific: Washington, Oregon, California, Alaska, Hawaii

[c]The Economic Research Service, U.S. Department of Agriculture, classifies rural counties by 1960-1990 poverty status and economic type. Counties not classified are urban counties and rural counties for which a classification could not be made.

SOURCE: Data from U.S. Census Bureau.

TABLE C-2 Comparison of Model Estimates and Other Procedurees with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Average Proportional Algebraic Difference for Counties in Each Category (in percent)

		Model			
Category	Counties (Number)	Log No. Under 21 (a)	Log No. Under 18 (b)	Log Rate Under 21 (c)	Log Rate Under 18 (d)
Census Division					
New England	67	4.1	4.5	6.6	7.1
Middle Atlantic	150	–5.9	–8.4	0.7	–1.0
East North Central	437	–3.6	–3.0	2.5	3.0
West North Central	618	–3.1	–0.6	0.5	2.3
South Atlantic	591	1.2	2.5	8.9	9.8
East South Central	364	–4.6	–3.0	0.5	1.3
West South Central	470	–7.6	–4.6	–4.0	–2.3
Mountain	281	0.6	5.4	7.2	10.4
Pacific	163	10.2	13.6	17.8	20.2
Metropolitan Status					
Central county of metropolitan area	493	0.6	–2.0	1.0	–0.6
Other metropolitan	254	–3.6	–0.8	11.6	13.7
Nonmetropolitan	2,394	–2.6	0.2	2.9	4.7
1990 Population Size					
under 7,500	525	–5.9	1.6	2.6	7.6
7,500-14,999	630	–1.0	3.0	5.7	8.4
15,000-24,999	524	–3.2	–1.8	2.1	3.2
25,000-49,999	620	–1.5	–0.7	4.2	4.6
50,000-99,999	384	–1.4	–3.3	2.5	1.2
100,000-249,999	259	–0.7	–3.4	1.5	–0.3
250,000 or more	199	1.0	0.4	1.3	1.1
1980 to 1990 Population Growth					
Decrease of more than 10.0%	444	–5.2	–1.0	–1.2	2.0
Decrease 0.1-10.0%	972	–3.3	–2.2	0.1	0.9
0.0-4.9%	547	–1.3	0.4	4.0	5.0
5.0-14.9%	620	–0.7	0.0	4.7	5.0
15.0-24.9%	260	4.0	3.8	10.6	10.1
25.0% or more	292	–4.1	2.3	9.8	14.0

Other Procedures			
Stable Shares (i)	Stable Shares in State (ii)	Stable Rates in State (iii)	Average of Census and (a) (iv)
45.6	7.0	8.6	20.2
28.8	–0.2	3.1	3.6
0.6	3.5	5.8	–4.6
18.7	21.0	15.9	3.7
28.6	10.2	11.9	14.5
19.5	0.4	0.3	5.0
–6.4	8.8	–0.2	–5.5
–3.4	30.5	22.6	2.6
–9.6	23.9	20.6	7.5
4.2	–0.2	2.2	0.8
16.2	7.0	20.9	11.7
13.2	15.0	9.9	3.6
30.3	42.0	25.9	9.2
16.3	17.5	12.2	6.1
9.0	6.8	4.5	1.1
6.0	3.1	5.3	2.2
3.1	–1.7	3.3	0.8
2.4	–2.5	2.8	0.8
7.9	2.9	6.5	4.5
29.0	36.9	17.5	3.7
11.6	10.1	3.0	–0.8
11.7	7.5	5.2	3.3
9.9	6.1	8.7	4.8
8.7	8.7	16.0	10.4
–4.0	4.3	23.8	12.6

continued on next page

TABLE C-2 Continued

Category	Counties (Number)	Model			
		Log No. Under 21 (a)	Log No. Under 18 (b)	Log Rate Under 21 (c)	Log Rate Under 18 (d)
Percent Poor School-Age Children, 1980					
Less than 9.4%	516	–4.1	–3.0	3.7	5.2
9.4-11.6%	524	–1.7	–0.2	2.4	3.6
11.7-14.1%	530	–2.0	–1.2	1.4	2.0
14.2-17.2%	523	–0.3	0.8	3.9	4.7
17.3-22.3%	519	–2.6	–1.2	1.9	2.6
22.4-53.0%	523	–2.3	3.2	6.3	9.3
Percent Hispanic, 1990					
0.0-0.9%	1,770	–3.2	–1.4	2.6	3.9
1.0-4.9%	847	1.0	3.1	7.1	8.3
5.0-9.9%	193	–0.6	0.7	2.2	3.3
10.0-24.9%	181	–5.7	–3.0	–2.9	–1.2
25.0-98.0%	150	–6.2	–3.3	–2.2	–0.3
Percent Black, 1990					
0.0-0.9%	1,446	–2.4	1.4	4.0	6.7
1.0-4.9%	615	–1.4	–2.1	3.1	2.4
5.0-9.9%	294	–2.4	–2.4	2.6	2.6
10.0-24.9%	381	–0.7	0.6	4.7	5.4
25.0-87.0%	405	–3.8	–2.7	0.0	0.9
Persistent Rural Poverty, 1960-1990					
Rural, not poor	1,740	–2.6	0.0	2.3	4.1
Rural, poor	535	–3.7	0.3	3.5	5.5
Not classified	866	–0.4	–1.1	5.2	4.8
Economic Type, Rural Counties					
Farming	556	–5.2	0.3	0.3	4.2
Mining	146	–8.6	–1.2	–1.7	2.2
Manufacturing	506	–3.8	–2.2	2.6	3.9
Government	243	5.8	5.1	11.8	10.5
Services	323	–2.1	–0.4	1.6	2.7
Nonspecialized	484	–2.8	–0.1	1.9	3.7
Not classified	883	–0.1	–0.8	5.4	5.1

Other Procedures			
Stable Shares (i)	Stable Shares in State (ii)	Stable Rates in State (iii)	Average of Census and (a) (iv)
1.9	2.9	8.1	–0.4
3.5	6.0	6.1	0.6
5.6	8.3	6.2	0.5
15.6	17.0	13.6	6.0
17.0	15.1	9.8	5.1
28.7	22.4	13.6	11.1
20.7	12.1	10.2	5.4
4.7	10.4	11.0	4.5
–0.6	15.4	10.2	1.0
–7.1	14.8	5.1	–3.5
–10.0	11.7	–1.2	–5.8
12.7	19.9	15.9	4.1
5.3	5.1	3.8	0.3
5.7	3.2	4.9	2.4
13.8	5.9	8.0	8.0
23.1	6.2	0.5	5.3
12.5	16.4	11.4	3.0
16.2	12.0	4.0	4.4
8.6	3.0	9.3	5.1
29.0	37.3	22.6	7.5
–2.4	11.9	3.3	–4.0
17.3	7.0	5.1	4.0
5.8	12.1	9.3	5.0
2.6	6.4	5.9	0.4
6.8	7.1	3.7	0.8
8.8	3.5	9.6	5.3

continued on next page

TABLE C-2 Continued

Category	Counties (Number)	Model: Log No. Under 21 (a)	Log No. Under 18 (b)	Log Rate Under 21 (c)	Log Rate Under 18 (d)
Percent Group Quarters Residents, 1990					
Less than 1.0%	545	–5.7	2.5	6.1	11.4
1.0-4.9%	2,187	–3.1	–0.6	1.7	3.7
5.0-9.9%	299	5.2	–0.6	6.7	1.7
10.0-41.0%	110	13.8	–5.0	11.5	–3.9
Status in CPS, 1989-1991					
In CPS sample	1,028	–0.9	–1.3	1.9	1.7
In CPS, no poor children 5-17	246	–1.3	1.0	9.9	11.6
Not in CPS sample	1,867	–3.0	0.2	3.1	5.2
Change in Poverty Rate for School-Age Children, 1980-1990					
Decrease of more than 3.0%	536	12.5	19.1	25.6	30.0
Decrease 0.1-3.0%	649	2.0	3.6	9.2	10.3
0.0-0.9%	272	–0.9	–0.1	4.9	5.4
1.0-3.4%	621	–3.7	–4.0	–0.3	–0.4
3.5-6.4%	532	–7.8	–7.7	–6.3	–6.2
6.5-38.0%	523	–15.5	–12.9	–13.8	–12.3

NOTE: See Notes to Table C-1.
SOURCE: Data from U.S. Census Bureau.

COMPARISONS FOR CANDIDATE MODELS AND OTHER ESTIMATION PROCEDURES

The four candidate models considered in Chapter 6 have the following designations: (a) log number model (under 21); (b) log number model (under 18); (c) log rate model (under 21); and (d) log rate model (under 18).[3] The four other

[3]The estimates from the four candidate models and the models considered in the first round of evaluations, listed below, are the final estimates for all counties, after the initial estimates from the county regression model are combined in a "shrinkage procedure" with direct CPS estimates for those counties with households in the CPS sample and raked for consistency with the estimates from the state model; see Chapter 4.

Other Procedures			
Stable Shares (i)	Stable Shares in State (ii)	Stable Rates in State (iii)	Average of Census and (a) (iv)
16.4	17.6	15.8	8.4
11.3	11.2	8.6	3.0
11.5	9.6	6.7	3.0
7.7	6.4	5.0	–0.7
7.9	2.8	4.4	2.7
20.5	11.2	19.0	11.3
13.2	17.2	11.2	3.5
71.8	65.8	61.7	41.4
28.1	19.2	20.6	13.9
9.5	9.8	9.3	3.5
–0.9	1.9	0.1	–4.2
–13.4	–8.2	–12.4	–12.6
–26.5	–18.6	–23.7	–20.9

procedures (see Chapter 6) are designated as follows: (i) stable shares; (ii) stable shares within state; (iii) stable rates within state (with conversion); and (iv) average of 1980 census estimates and estimates for 1989 from the log number (under 21) model (a).

Table 6-3 presents the overall measures of average absolute difference (measure 1) and average proportional absolute difference (measure 2) between the estimates from the four candidate models and four procedures and the estimates from the census. Table 6-4 presents the category algebraic differences (measure 3) for the four candidate models and procedures (i) and (iv). Table C-1 is identical to Table 6-4 except that it also includes results for procedures (ii) and (iii). Table C-2 presents the category average proportional algebraic differences for the four candidate models and the four procedures. For reasons given in Chapter 6, the 1990 census estimates used in these comparisons are ratio-adjusted by a constant factor to equal the CPS national estimate of poor school-age children in 1989.

The findings from these evaluations are discussed in Chapter 6. The additional detail in Tables C-1 and C-2 is presented without commentary.

COMPARISONS FOR THE SINGLE-EQUATION MODELS CONSIDERED IN THE FIRST ROUND OF EVALUATIONS

Six single-equation models were considered in the first round of evaluations (see Chapter 5). For this appendix these models are labeled as follows: (C.1) log number model (under 21) (model (a) of the candidate models); (C.2) log number model (under 18) (model (b) of the candidate models); (C.3) log number model (under 21) with fixed state effects; (C.4) log rate model (under 21) (model (c) of the candidate models); (C.5) rate model (under 21, variables not transformed); and (C.6) hybrid log rate-number model (under 21).[4] Also included are comparisons for a variant of each of the three rate models—C.4a, C.5a, and C.6a, respectively—in which 1990 census population figures instead of estimates from the Census Bureau's population estimates program are used to convert the estimated proportions of poor school-age children from each rate model to estimated numbers.

For the first round of evaluations the census estimates were not ratio-adjusted to make the census national estimate of poor school-age children in 1989 equal to the corresponding CPS total for 1989, unlike the situation with the evaluations of the candidate models and other procedures described above. Thus, the results of the first round of evaluations given in Tables C-3 to C-5 cannot be directly compared with those for the later round. However, knowing that the ratio-adjustment increased the census estimates by about 5 percent, it could be possible to make some rough comparisons.

Overall Differences

Table C-3 presents the average absolute difference (measure 1) and the average proportional absolute difference (measure 2) between model estimates and 1990 census estimates of the number of poor school-age children in 1989 for the six single-equation models, C.1-C.6, that were included in the first round of county model evaluations. It also shows the two absolute difference measures for the variant of the three rate models, C.4a, C.5a, and C.6a, in which 1990 census population figures instead of estimates from the Census Bureau's population

[4]The "under 21" designation is retained in the discussion only for the log number model, C.1, to distinguish it from model C.2.

TABLE C-3 Comparison of First-Round Model Estimates with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989

Model		Average Absolute Difference	Average Proportional Absolute Difference, in Percent
C.1	Log number model (under 21)	284	15.7
C.2	Log number model (under 18)	284	17.1
C.3	Log number model (under 21), with fixed state effects	289	17.4
C.4	Log rate model (under 21), rates converted to numbers with 1990 population estimates	285	18.9
C.4a	Log rate model (under 21), rates converted to numbers with 1990 census estimates	263	17.9
C.5	Rate model (under 21), untransformed, rates converted to numbers with 1990 population estimates	325	20.0
C.5a	Rate model (under 21), untransformed, rates converted to numbers with 1990 census estimates	299	18.8
C.6	Hybrid log rate-number model (under 21), rates converted to numbers with 1990 population estimates	298	17.1
C.6a	Hybrid log rate-number model (under 21), rates converted to numbers with 1990 census estimates	270	15.3

NOTE: See text for definitions of models and measures.
SOURCE: Data from U.S. Census Bureau.

estimates program are used to convert estimated proportions to estimated numbers of poor school-age children.

For models C.1, C.2, C.3, C.4, C.5, and C.6, the average absolute difference ranges from 284 to 325, or 11-13 percent of the average number of poor school-age children per county for 1989 (about 2,500 children). For these six models, the average proportional absolute difference ranges from 15.7 to 20.0 percent. The log number (under 21) model (C.1) performs best; it has the lowest average proportional absolute difference and is tied with the log number (under 18) model

(C.2) for the lowest average absolute difference. The rate model (D.5) performs worst; it has the largest differences on both measures.

Because the 1990 census estimates used in the comparisons for models C.1-C.6 are not ratio-adjusted to the CPS national estimate of poor school-age children in 1989, the absolute difference measures in Table C-3 are about 5 percent higher than they would be if the ratio-adjustment had been made.[5] For an evaluation of overall differences, controlling the 1990 census estimates to the CPS national estimate does not affect comparisons across models. However, for evaluation of category differences, there could be an effect.

Use of 1990 Population Estimates

For rate models, it is necessary to use population estimates of the number of school-age children to convert estimated proportions to estimated numbers of poor school-age children. The population estimates themselves differ from 1990 census figures (see Chapter 8). The use of 1990 population estimates instead of 1990 census figures to convert estimated proportions from the three rate models to estimated numbers increases the average absolute difference in the estimated number of poor school-age children by 8-10 percent and increases the average proportional absolute difference by about 6 percent for the log rate and rate models and 12 percent for the hybrid log rate-number model. (Compare the measures in Table C-3 for model C.4 and C.4a, for C.5 and C.5a, and for C.6 and C.6a.)

Differences by Categories of Counties

Tables C-4 and C-5 (on pages 214-225) show the category algebraic differences (measure 3) and the category average proportional algebraic differences (measure 4), respectively, between model estimates and 1990 census estimates of the number of poor school-age children in 1989 for the six single-equation models that were considered in the first round of county model evaluations and the variant of the three rate models. The discussion considers models C.1-C.6.

Census Division The category algebraic differences in the predicted number of poor school-age children categorized by census division (measure 3, Table C-4) are the same for all of the models because they are raked to the same set of state estimates. They vary widely by census division. In particular, all of the models overpredict the number of poor school-age children for counties in the Mountain Division and, especially, the Pacific Division relative to other counties.

[5]Comparing Tables C-3 and 6-3, the average absolute differences for models C.1, C.2, and C.4 from Table C-3 are 4 to 6 percent higher than the corresponding differences for models (a), (b), and (c) from Table 6-3; the average proportional absolute differences are 2 to 8 percent higher.

The proportional category differences (measure 4, Table C-5) vary even more widely across divisions than do the category differences. For the Pacific Division, the proportional category difference is 1.3 to 2 times the category difference (16-26% versus 12%), indicating that the overprediction is more pronounced for smaller counties than larger counties in that geographic area.[6] Further investigation is required to determine the reasons for the variations across divisions, which could include sampling variability in the CPS for 1989 or a specification problem in the state model (see Chapter 6).

Metropolitan Status The category differences and proportional category differences in the predicted number of poor school-age children vary somewhat for counties categorized by metropolitan status. There is no consistent pattern across models: for example, the log number (under 21) model (C.1) overpredicts the number of poor school-age children in central counties of metropolitan areas relative to other counties, while the log rate model (C.4) overpredicts the number of poor school-age children in "other metropolitan" counties relative to central counties or counties in nonmetropolitan areas.

1990 Population Size The category differences in the predicted number of poor school-age children (Table C-4) show a systematic tendency for the log number (under 21) model (C.1) and the hybrid log rate-number model (C.6) to overpredict the number of poor school-age children for larger size counties relative to smaller size counties. The proportional category differences (Table C-5) show somewhat less variation. A statistical test established that the variations in the proportional differences for categories of counties classified by population size were significant for model C.6, but not for model C.1. However, the test used was not sensitive to monotonic patterns—for example, an increasing rate of overprediction by county size. (The test was not performed for the category differences, measure 3.)

Population Growth from 1980-1990 The category differences and proportional category differences in the predicted number of poor school-age children show a tendency for most models to overpredict the number of poor school-age children in counties with larger rates of population increase from 1980 to 1990 relative to counties with smaller increases or with decreases.[7] However, the extent of overprediction does not increase monotonically. In particular, most

[6]The proportional category differences differ somewhat across models because they are calculated relative to each county's 1990 census estimated number of poor school-age children before being summed.

[7]A statistical test established that the variations in the proportional category differences for categories of counties classified by population growth rate were significant for three of the four models tested: C.1, C.2, and C.3, but not C.6.

models underpredict the number of poor school-age children for counties with the largest population increases (25% or more) relative to counties with the next largest increases (15-25%). In contrast to the pattern shown by other models, the log number model with fixed state effects (C.3) tends to overpredict the number of poor school-age children for counties that experienced a large population decrease relative to other counties.

Percentage of Poor School-Age Children, 1980 Census The category differences and proportional category differences in the predicted number of poor school-age children show relatively little variation for most models for counties categorized by their proportion of poor school-age children in 1979. The exception is the log number model with fixed state effects (C.3), which overpredicts the number of poor school-age children for counties that had a higher proportion of such children in 1979 relative to counties with a lower proportion. The variation in the proportional category differences (Table C-5) for counties defined by their 1979 proportion of poor school-age children is statistically significant for this model.

Percentage of Hispanic Population in 1990 The category differences in the predicted number of poor school-age children (Table C-4) show a tendency for most models to overpredict the number of poor school-age children for counties with larger proportions of Hispanics relative to other counties. This pattern is particularly pronounced for the log number (under 21 and under 18) models (C.1, C.2). The proportional category differences (Table C-5) tend to show the opposite pattern, in which the number of poor school-age children is overpredicted for counties with *smaller* proportions of Hispanics relative to other counties. The variations in the proportional category differences for counties characterized by percentage of Hispanic population are statistically significant for all models with this pattern that were tested. The differences in the patterns for the two measures may occur because the models behave differently for small counties with many Hispanics (primarily rural border counties) than for large counties (cities).

Percentage of Black Population in 1990 The category differences in the predicted number of poor school-age children (Table C-4) show a slight tendency for the log rate and rate models (C.4, C.5) to overpredict the number of poor school-age children for counties with smaller proportions of blacks relative to other counties. The proportional category differences (Table C-5) show little variation for any of the models for counties characterized by percentage of black population in 1990.

Persistent Rural Poverty, 1960-1990 The category differences in the predicted number of poor school-age children (Table C-4) vary little for most models for counties characterized as rural and persistently poor, rural and not persistently poor, and not classified (urban counties and rural counties for which a classification could not be made). However, the log number (under 21) model (C.1) underpredicts the number of poor school-age children for rural counties relative to not classified counties. Also, the hybrid log rate-number model (C.6) underpredicts the number of poor school-age children for rural counties, whether or not they are persistently poor, relative to not classified counties. This pattern, which appears for both category difference measures, is statistically significant for the proportional category difference measure (Table C-5).

Economic Type, Rural Counties The category differences and proportional category differences in the predicted number of poor school-age children vary for all models for rural counties categorized by their principal economic activity. In particular, all of the models overpredict the number of poor school-age children in rural counties that have a large government presence relative to other types of rural counties.

Percentage of Group Quarters Residents in 1990 The category differences and proportional category differences in the predicted number of poor school-age children show that the log number (under 21) model (C.1), log number model with fixed state effects (C.3), and log rate model (C.4) tend to overpredict the number of poor school-age children in counties with larger percentages of group quarters residents relative to other counties. The pattern is particularly strong for model C.1. As discussed in Chapter 6, the replacement of the population under age 21 as a predictor variable in model C.1 by the population under 18 in model C.2 removed this pattern.

Status in CPS, 1989-1991 The category differences and proportional category differences in the predicted number of poor school-age children are similar in most models for counties categorized by their representation in the CPS sample. The log rate model (C.4) overpredicts the number of poor school-age children in counties with CPS sampled households, none of which contain poor school-age children (and thereby are excluded from the sample for estimating the model),[8] relative to other counties. The hybrid log rate-number model (C.6) somewhat overpredicts the number of poor school-age children in counties with CPS sampled households relative to counties with no CPS sampled households.

[8]The only model that uses these counties in the estimation is the rate model for which the variables are untransformed (C.5).

TABLE C-4 Comparison of First-Round Model Estimates with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Algebraic Difference by Category of County (in percent)

Category	Model: Log Number Under 21 C.1	Log Number Under 18 C.2	Log Number Under 21, Fixed State Effects C.3
Census Division[a]			
New England	1.9	1.9	1.9
Middle Atlantic	2.0	2.0	2.0
East North Central	4.7	4.7	4.7
West North Central	6.8	6.8	6.8
South Atlantic	5.5	5.5	5.5
East South Central	0.3	0.3	0.3
West South Central	2.1	2.1	2.1
Mountain	9.4	9.4	9.4
Pacific	11.8	11.8	11.8
Metropolitan Status			
Central county of metropolitan area	7.4	6.7	6.6
Other metropolitan	–2.0	–0.3	–3.9
Nonmetropolitan	0.5	2.0	2.8
1990 Population Size			
under 7,500	–4.5	2.5	4.7
7,500-14,999	0.4	5.5	6.0
15,000-24,999	–0.4	2.3	2.8
25,000-49,999	0.5	1.8	1.9
50,000-99,999	1.2	–0.4	–0.1
100,000-249,999	3.1	0.4	1.1
250,000 or more	8.4	8.3	7.9
1980 to 1990 Population Growth			
Decrease of more than 10.0%	3.0	5.6	9.0
Decrease 0.1-10.0%	4.3	4.4	5.9
0.0-4.9%	2.0	2.0	2.5
5.0-14.9%	5.0	3.8	3.8
15.0-24.9%	13.1	11.1	10.9
25.0% or more	0.7	3.5	–0.5

Log Rate Under 21		Rate Under 21		Log Hybrid Rate-Number Under 21	
C.4	C.4a	C.5	C.5a	C.6	C.6a
1.9	1.9	1.9	1.9	1.9	1.9
2.0	2.0	2.0	2.0	2.0	2.0
4.7	4.7	4.7	4.7	4.7	4.7
6.8	6.8	6.8	6.8	6.8	6.8
5.5	5.5	5.5	5.5	5.5	5.5
0.3	0.3	0.3	0.3	0.3	0.3
2.1	2.1	2.1	2.1	2.1	2.1
9.4	9.4	9.4	9.4	9.4	9.4
11.8	11.8	11.8	11.8	11.8	11.8
4.8	4.5	4.8	4.5	7.7	7.4
10.2	7.5	9.7	7.0	2.5	–0.1
4.6	5.8	4.6	5.8	–0.9	0.2
3.0	4.4	5.6	7.2	–6.6	–5.3
7.6	8.6	7.7	8.7	–0.9	0.0
5.3	6.4	5.2	6.3	–1.5	–0.4
5.6	6.1	5.5	6.0	0.3	0.7
3.6	3.9	3.8	4.0	0.3	0.6
3.0	3.1	1.7	1.8	2.1	2.2
5.5	5.0	5.7	5.3	9.2	8.8
1.3	1.9	2.4	3.0	2.4	3.0
2.9	3.0	3.1	3.2	3.9	4.0
1.6	2.3	1.2	1.9	1.3	2.0
5.2	5.1	5.6	5.6	4.2	4.2
10.7	9.9	10.9	10.0	12.6	11.7
6.7	6.6	5.8	5.6	4.1	3.9

continued on next page

TABLE C-4 Continued

Category	Model: Log Number Under 21 C.1	Log Number Under 18 C.2	Log Number Under 21, Fixed State Effects C.3
Percent Poor School-Age Children, 1980			
Less than 9.4%	0.8	0.2	–1.0
9.4-11.6%	4.4	3.9	3.3
11.7-14.1%	8.8	7.3	7.0
14.2-17.2%	5.8	6.2	5.2
17.3-22.3%	6.8	6.7	8.5
22.4-53.0%	2.6	5.7	7.7
Percent Hispanic, 1990			
0.0-0.9%	1.4	1.4	2.3
1.0-4.9%	5.5	5.0	4.7
5.0-9.9%	3.5	4.3	3.3
10.0-24.9%	7.3	6.8	7.4
25.0-98.0%	9.0	9.8	8.5
Percent Black, 1990			
0.0-0.9%	3.6	5.2	5.3
1.0-4.9%	4.2	2.8	2.9
5.0-9.9%	1.9	2.4	1.5
10.0-24.9%	7.0	6.2	5.7
25.0-87.0%	6.0	6.7	7.9
Persistent Rural Poverty, 1960-1990[b]			
Rural, not poor	0.8	1.0	1.4
Rural, poor	–0.3	2.7	5.2
Not classified	6.7	6.2	5.8
Economic Type, Rural Counties[b]			
Farming	–0.8	2.4	7.0
Mining	–6.3	–0.4	–4.0
Manufacturing	–1.6	–1.2	0.4
Government	7.2	3.6	8.7
Services	0.8	1.8	1.1
Nonspecialized	1.0	3.9	3.4
Not classified	6.7	6.2	5.8

Log Rate Under 21		Rate Under 21		Log Hybrid Rate-Number Under 21	
C.4	C.4a	C.5	C.5a	C.6	C.6a
4.9	1.7	5.6	2.3	5.6	2.3
3.2	3.0	4.4	4.2	5.6	5.3
6.8	6.9	6.2	6.4	7.5	7.6
3.7	6.7	2.8	5.7	2.7	5.7
5.3	5.8	4.4	4.8	5.0	5.6
6.3	6.8	6.2	6.7	2.1	2.7
3.3	3.1	3.1	3.0	1.6	1.4
5.4	5.1	5.1	4.8	5.6	5.3
3.8	3.4	5.0	4.7	4.4	3.9
5.7	5.1	7.2	6.4	8.2	7.6
7.2	8.9	5.9	7.7	6.9	8.6
9.0	9.1	8.6	8.7	4.3	4.3
6.3	5.6	6.9	6.1	4.8	4.0
4.2	3.6	4.1	3.6	4.3	3.6
3.9	3.8	3.9	3.8	6.5	6.3
3.1	4.2	2.9	4.1	4.2	5.5
3.6	5.4	3.6	5.3	–1.1	0.5
5.7	5.9	5.7	5.8	–1.4	–1.2
5.2	4.8	5.2	4.8	7.2	6.9
3.3	5.2	5.0	6.9	–3.9	–2.1
–1.7	1.5	–1.4	1.8	–6.0	–3.1
3.2	3.1	3.4	3.3	–1.5	–1.7
11.6	11.7	9.7	9.7	1.9	2.0
3.1	4.8	3.1	4.8	–0.5	1.2
4.8	6.8	4.4	6.3	–0.4	1.4
5.2	4.8	5.3	4.8	7.3	6.9

continued on next page

TABLE C-4 Continued

Category	Model: Log Number Under 21 C.1	Log Number Under 18 C.2	Log Number Under 21, Fixed State Effects C.3
Percent Group Quarters Residents, 1990			
Less than 1.0%	–2.1	2.1	–0.5
1.0-4.9%	5.2	5.7	5.4
5.0-9.9%	7.4	0.3	5.0
10.0-41.0%	19.9	1.6	11.9
Status in CPS, 1989-1991			
In CPS sample	6.4	5.9	5.8
In CPS, no poor children 5-17	2.2	3.0	0.8
Not in CPS sample	0.6	2.0	2.8

NOTES: See text for definitions of models and measures. 3,141 counties are assigned to a category for most characteristics; 3,135 counties are assigned to a category for 1980-1990 population growth and 1980 percentage of poor school-age children; 3,133 counties are assigned to a category for 1980-1990 percentage change in poverty rate for school-age children; see Table C-1 for number of counties in each category.

[a]Census division states:
New England: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
Middle Atlantic: New York, New Jersey, Pennsylvania
East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin
West North Central: Missouri, Minnesota, Iowa, North Dakota, South Dakota, Nebraska, Kansas

Log Rate Under 21		Rate Under 21		Log Hybrid Rate-Number Under 21	
C.4	C.4a	C.5	C.5a	C.6	C.6a
7.0	4.9	8.9	6.7	2.7	0.6
4.6	4.6	4.7	4.7	5.7	5.7
5.5	7.3	4.0	5.8	0.3	2.0
12.7	17.4	5.0	9.3	0.6	4.7
4.7	4.4	5.3	5.0	6.8	6.6
12.6	10.2	–1.0	–2.9	5.3	3.0
4.8	6.2	5.0	6.3	–1.4	–0.1

South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
East South Central: Kentucky, Tennessee, Alabama, Mississippi
West South Central: Arkansas, Louisiana, Oklahoma, Texas
Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada
Pacific: Washington, Oregon, California, Alaska, Hawaii

[b]The Economic Research Service, U.S. Department of Agriculture, classifies rural counties by 1960-1990 poverty status and economic type. Counties not classified are urban counties and rural counties for which a classification could not be made.

SOURCE: Data from U.S. Census Bureau.

TABLE C-5 Comparison of First-Round Model Estimates with 1990 Census County Estimates of the Number of Poor School-Age Children in 1989: Average Proportional Algebraic Difference for Counties in Each Category (in percent)

	Model		
Category	Log Number Under 21 C.1	Log Number Under 18 C.2	Log Number Under 21, Fixed State Effects C.3
Census Division			
New England	9.3	9.7	8.1
Middle Atlantic	–1.2	–3.9	–3.5
East North Central	1.2	1.8	2.2
West North Central	1.7	4.4	7.4
South Atlantic	6.2	7.6	8.1
East South Central	0.1	1.8	0.9
West South Central	–3.0	0.1	0.2
Mountain	5.6	10.6	12.2
Pacific	15.6	19.2	19.2
Metropolitan Status			
Central county of metropolitan area	5.6	2.9	3.5
Other metropolitan	1.1	4.1	–0.1
Nonmetropolitan	2.2	5.1	6.5
1990 Population Size			
under 7,500	–1.3	6.6	9.9
7,500-14,999	3.9	8.1	9.3
15,000-24,999	1.6	3.0	4.2
25,000-49,999	3.4	4.2	3.7
50,000-99,999	3.4	1.5	1.0
100,000-249,999	4.2	1.4	1.4
250,000 or more	5.9	5.4	5.0
1980 to 1990 Population Growth			
Decrease of more than 10.0%	–0.5	3.9	10.5
Decrease 0.1-10.0%	1.5	2.6	5.5
0.0-4.9%	3.6	5.3	5.1
5.0-14.9%	4.2	4.9	4.1
15.0-24.9%	9.2	9.0	7.5
25.0% or more	0.7	7.3	–0.3

Log Rate Under 21		Rate Under 21		Log Hybrid Rate-Number Under 21	
C.4	C.4a	C.5	C.5a	C.6	C.6a
11.9	13.1	10.9	12.2	8.3	9.4
5.7	4.1	4.2	2.8	–1.2	–2.6
7.5	8.5	6.4	7.4	–0.1	0.7
5.4	7.3	6.1	8.0	–0.2	1.6
14.3	12.6	14.5	12.8	7.7	6.1
5.4	4.8	5.3	4.6	0.7	0.0
0.7	3.3	1.8	4.3	–6.7	–4.4
12.5	14.6	17.0	19.3	3.9	5.7
23.7	23.8	25.6	25.8	15.6	15.7
6.0	4.9	5.0	4.0	6.1	4.9
17.1	13.3	16.1	12.4	6.8	3.4
7.9	9.4	9.0	10.5	0.3	1.6
7.7	9.2	12.7	14.2	–3.5	–2.3
10.9	12.3	11.5	12.8	2.2	3.4
7.2	8.2	6.9	8.0	0.1	1.1
9.3	10.1	8.8	9.6	2.8	3.5
7.5	7.3	7.3	7.0	3.1	2.8
6.6	6.0	3.3	2.9	4.0	3.4
6.3	4.4	7.3	5.5	8.7	6.9
3.7	3.9	7.9	8.0	–1.7	–1.5
5.0	6.4	5.4	6.8	–0.7	0.6
9.2	9.9	8.2	8.9	2.7	3.4
9.9	10.2	9.6	10.0	3.4	3.6
16.0	15.6	15.6	15.2	8.0	7.6
15.2	15.7	16.2	16.9	3.9	4.1

continued on next page

TABLE C-5 Continued

Category	Model: Log Number Under 21 C.1	Log Number Under 18 C.2	Log Number Under 21, Fixed State Effects C.3
Percent Poor School-Age Children, 1980			
Less than 9.4%	0.6	1.8	–1.3
9.4-11.6%	3.2	4.8	3.5
11.7-14.1%	2.9	3.6	4.3
14.2-17.2%	4.6	5.8	8.1
17.3-22.3%	2.2	3.7	6.9
22.4-53.0%	2.5	8.3	11.2
Percent Hispanic, 1990			
0.0-0.9%	1.6	3.5	5.1
1.0-4.9%	6.0	8.2	6.7
5.0-9.9%	4.3	5.7	6.4
10.0-24.9%	–1.1	1.8	3.1
25.0-98.0%	–1.5	1.5	4.7
Percent Black, 1990			
0.0-0.9%	2.4	6.5	7.3
1.0-4.9%	3.5	2.8	3.5
5.0-9.9%	2.4	2.4	1.8
10.0-24.9%	4.2	5.6	4.5
25.0-87.0%	0.9	2.1	5.6
Persistent Rural Poverty, 1960-1990			
Rural, not poor	2.2	4.9	6.1
Rural, poor	1.0	5.3	7.7
Not classified	4.5	3.8	2.9
Economic Type, Rural Counties			
Farming	–0.5	5.3	9.9
Mining	–4.1	3.7	0.7
Manufacturing	1.0	2.7	3.5
Government	11.0	10.3	13.2
Services	2.7	4.5	4.3
Nonspecialized	2.0	4.9	4.8
Not classified	4.8	4.1	3.3

	Log Rate Under 21		Rate Under 21		Log Hybrid Rate-Number Under 21	
	C.4	C.4a	C.5	C.5a	C.6	C.6a
	8.9	7.5	7.2	5.9	3.8	2.5
	7.5	9.0	8.5	10.2	2.3	3.7
	6.4	7.7	7.3	8.6	0.9	2.2
	9.1	11.1	10.2	12.3	1.4	3.2
	7.0	8.1	7.5	8.6	–0.6	0.4
	11.5	10.6	13.0	11.0	2.4	1.5
	7.7	7.6	8.4	8.2	1.5	1.4
	12.4	12.9	12.1	12.7	5.6	6.1
	7.2	10.0	9.3	12.4	0.6	3.0
	1.9	5.4	3.3	6.9	–6.0	–2.9
	2.6	7.4	3.8	8.7	–7.7	–3.5
	9.2	10.4	10.4	11.7	1.5	2.6
	8.2	8.7	8.0	8.7	2.0	2.4
	7.7	6.6	7.9	6.9	3.1	2.1
	9.9	9.8	9.0	9.0	3.9	3.7
	5.0	5.4	5.9	6.2	–1.0	–0.5
	7.3	9.4	8.7	10.8	0.1	1.9
	8.6	8.3	8.4	8.1	0.0	–0.2
	10.3	8.7	9.7	8.2	6.0	4.5
	5.3	7.6	9.3	11.6	–3.5	–1.3
	3.1	8.6	4.5	10.3	–6.7	–2.0
	7.6	7.6	7.5	7.4	1.2	1.1
	17.3	17.2	15.0	14.8	7.2	7.0
	6.6	8.2	7.9	9.6	0.8	2.3
	7.0	8.7	6.9	8.6	0.3	2.0
	10.6	9.0	10.1	8.6	6.3	4.8

continued on next page

TABLE C-5 Continued

	Model		
Category	Log Number Under 21 C.1	Log Number Under 18 C.2	Log Number Under 21, Fixed State Effects C.3
Percent Group Quarters Residents, 1990			
Less than 1.0%	–1.1	7.5	3.8
1.0-4.9%	1.7	4.3	5.0
5.0-9.9%	10.4	4.3	9.5
10.0-41.0%	19.4	–0.3	12.4
Status in CPS, 1989-1991			
In CPS sample	4.0	3.6	3.9
In CPS, no poor children 5-17	3.6	6.0	3.1
Not in CPS sample	1.8	5.1	6.7

NOTE: See notes to Table C-4.

SOURCE: Data from U.S. Census Bureau.

Summary of Category Differences

Three of the eleven characteristics examined show no pronounced patterns of overprediction or underprediction of the number of poor school-age children for any of the models:

—percentage of poor school-age children from the 1980 census;
—percentage of black population in 1990; and
—persistent rural poverty from 1960 to 1990.

Four characteristics show patterns for all or all but one model in which some categories of counties are over(under)predicted relative to other counties:

—census geographic division;
—percentage of change in population from 1980 to 1990 (population growth);
—percentage of Hispanic population in 1990; and
—economic type, for rural counties.

Log Rate Under 21		Rate Under 21		Log Hybrid Rate-Number Under 21	
C.4	C.4a	C.5	C.5a	C.6	C.6a
11.3	10.7	13.7	13.2	1.8	1.1
6.7	7.4	7.4	8.1	1.5	2.1
11.9	14.2	11.3	13.5	3.3	5.3
17.0	19.0	9.9	11.8	2.0	3.8
7.0	6.6	8.9	8.6	4.3	3.9
15.4	13.9	4.6	3.5	5.8	4.5
8.2	9.7	9.5	11.0	–0.3	1.1

The remaining four characteristics exhibit mixed patterns, in which some models give evidence of over(under)prediction for counties in some categories and other models do not:

—metropolitan status of county;
—1990 population size;
—percentage of group quarters residents in 1990; and
—status in CPS sample.

Of these four characteristics, over(under)prediction for those models in which it occurs is most pronounced for population size and percentage of group quarters residents.

Overall, there is no clearly best or worst model in terms of differences from the 1990 census estimates for categories of counties. Each model exhibits strengths and weaknesses (keeping in mind that the analysis is based on a single evaluation). On balance, the log number (under 18) model (C.2) performs somewhat better than the other models.

APPENDIX D

Use of School Lunch Data in New York State for the Estimation of School-Age Children in Poverty: An Analysis

James H. Wyckoff and Frank Papa

This analysis uses data from the National School Lunch Program in New York State as an alternative to census data in estimating the number of poor children (aged 5-17) for use in the allocation of Title I funds to school districts. This analysis considers two uses of poverty estimates in the Title I allocations. First, for the purpose of estimating the number of school-age children who are in poor families in 1989, we compare estimates from using school lunch data for 1990 with estimates from the Census Bureau's constant-share method that is based on 1980 census data. Second, we examine the sensitivity of various methods in estimating the 15 percent threshold for concentration grants. In conclusion, we examine some of the difficulties we encountered in attempting to use school lunch data for this purpose. Although this analysis may provide some interesting insights to some evaluation questions, it only reflects the experience in one state; other states may well differ in critical ways that would lead outcomes to change as well.

The data for this analysis cover public schools and come from the New York State Education Department Report 325 for February 1990, printed on July 10, 1992. The 325 Report is an accounting of the number of eligible applicants for free and reduced-price school lunches by school. Our data include all public school reports.[1]

[1]Some of the state's 3,279 public schools did not send reports to the New York State Education Department: most of those 389 schools did not operate a school lunch program. Those observations are treated as zeros in the analysis.

Reports from private schools are also available but they have not been included in this analysis: 804 private schools reported 42,828 free and reduced-price school lunch applicants in February 1990. This number represents about 12 percent of all school lunch applicants.

ESTIMATES OF POOR CHILDREN

The school lunch method for estimating the number of poor children in each school district is conceptually similar to the census constant-share method. County totals of poor children are allocated to specific school districts on the basis of an estimate of the ratio of poor children in the district to the county total. The school lunch ratio is computed by the ratio of free (or free and reduced-price) school lunch applicants in a school district to those in the county. This ratio is then multiplied by the total number of poor school-age children in the county (from the 1990 census) to arrive at the school district estimate. When districts cross county boundaries, the district is assigned to the county in which the school district administrative office is located.[2] In summary:

$$\tilde{Y}j' = \frac{SL_{ij}^{90}}{SL_{i.}^{90}} Cen_i^{90} \quad , \tag{1}$$

where:

$\tilde{Y}j'$ is the school lunch estimate of poor school-age children in school district j,
SL_{ij}^{90} is the number of school lunch applicants in county i, school district j in 1990,
$SL_{i.}^{90}$ is the number of school lunch applicants in county i in 1990, and
CEN_i^{90} is the 1990 census estimate of poor school-age children in county i.

The evaluation below compares these estimates of poor school-age children to those estimated using the census constant-share method, which applies the 1980 census shares of poor school-age children for school districts (or parts of school districts) within counties to the 1990 census county estimates of poor school-age children (synthetic method (2) in Chapter 7). Mean algebraic and absolute percentage errors are estimated for each method by using the 1990 census totals for school districts as "truth." Tables D-1 to D-3 summarize these results.

Table D-1 illustrates the distribution of the algebraic percentage errors, unweighted and when each district is weighted by the number of school-age

[2]We also computed estimates by employing school-level data to form county pieces when schools of a district are located in more than one county. Roughly 35 percent of the districts cross county boundaries. This estimation method produces estimates that are very close to the method that does not account for the county pieces. As a result, we present only the results that assign a whole district to the county of the district's administrative office.

TABLE D-1 Distribution of Algebraic Percentage Errors for Children Aged 5-17 in Families in Poverty, Various Models, Unweighted and Weighted, New York State School Districts in Evaluation Universe, 1990 (N = 623), in percent

Distribution of Algebraic Percentage Errors	Census Constant 1980 Share	Free Lunch	Free and Reduced-Price Lunch
Unweighted			
Mean	31.2	7.1	14.0
Less than –40.0%	10.3	20.4	17.7
–40.0 to –20.0%	11.4	12.4	12.8
–19.9 to –10.0%	10.6	9.5	9.1
–9.9 to –0.1%	9.0	9.5	10.1
0.0 to 9.9%	10.6	12.0	9.8
10.0 to 19.9%	7.7	6.6	7.5
20.0 to 39.9%	11.9	11.9	10.3
40.0% and more	28.6	17.8	22.6
Weighted by Related Children Age 5-17 in Poverty, 1990 Census			
Mean	0.8	1.6	1.3
Less than –40.0%	5.0	8.1	6.8
–40.0 to –20.0%	16.0	11.0	13.7
–19.9 to –10.0%	17.0	10.0	12.1
–9.9 to –0.1%	28.6	10.8	29.8
0.0 to 9.9%	7.2	33.8	11.8
10.0 to 19.9%	8.1	8.1	5.0
20.0 to 39.9%	8.1	7.9	8.9
40% and more	10.1	10.3	11.9

NOTES: The census constant 1980 share estimates are calculated as described in Chapter 7 (within-county shares method (2)). The school lunch estimates are formed by multiplying the 1990 census estimates of related children aged 5-17 in families in poverty for the county by the school district's share of the county's free (free and reduced-price) lunch participants. The mean unweighted algebraic percentage error is the sum over all school districts of the algebraic difference between the estimate of poor school-age children from a model and the 1990 census estimate as a proportion of the census estimate for each district, divided by the number of districts. The weighted mean weights each difference by the census number of poor school-age children in the district.

children from families in poverty. Each of the methods results in estimates with some very large errors. For example, consider the weighted results. All three methods have at least 15 percent of the districts with errors of at least 40 percent. This pattern is also illustrated in Table D-2, which shows unweighted estimates broken down by various school district characteristics. Regardless of method, the errors are very large on average and in most categories.

Weighting by the number of poor school-age children in 1990 substantially reduces the percentage errors across all methods, as shown in Table D-3. This approach yields results that are quite similar across all three models. Mean algebraic percentage errors are relatively small; however, as one would expect, mean absolute percentage errors are much larger. Most of the patterns of errors with respect to school district attributes are as would be expected. For example, school districts with small total population have larger errors than districts with larger populations.

An important result of this analysis is that even after some effort in data preparation, the school lunch method is still not meaningfully better than the census constant-share method. At least in New York State it does not appear that using school lunch data results in significant gains in estimating school-age children from poor families.

ESTIMATES OF THE CONCENTRATION GRANT THRESHOLD

Eligibility for Title I concentration grants is based on having a school-age poverty rate of at least 15 percent or at least 6,500 poor children.[3] Current Title I allocations employ a two-stage eligibility criterion. A district must be in a county that meets the 15 percent (or 6,500) rule, and the district itself must meet that criterion. Under the proposed direct allocation system, grants will be made directly to districts and, as such, eligibility will be determined solely with regard to district poverty rates, without regard to county poverty rates. The proposed direct allocation method also permits states to aggregate the allocations to districts that have total population of less than 20,000 and reallocate this total based on alternative data, such as those from the National School Lunch Program. It is of interest to examine eligibility for concentration grants in those districts with less than 20,000 population under three different scenarios: the current two-stage process, the direct allocation process to districts without controls, and direct allocations when school district poverty estimates must sum to the census county totals. We examine how concentration grants eligibility differs under these circumstances when school lunch data are used rather than census constant-share estimates, using 1990 census ratio-adjusted counts as the measure of truth.

[3]Children eligible for Title I are not limited to school-age children from poor families (see Chapter 2). However, for the purpose of this analysis, which is to examine the census constant-share estimates of school-age children from poor families, eligibility is so characterized.

TABLE D-2 Mean Absolute and Algebraic Percentage Errors for Children Aged 5-17 in Families in Poverty, Various Methods, New York State School Districts in Evaluation Universe, 1990, Unweighted, in percent

		Census Constant 1980 Share	
Category	Percent of Districts (N = 623)	Mean Absolute % Error	Mean Algebraic % Error
Total	100.0	53.4	31.2
1990 School District Population			
Under 2,500	11.9	66.3	34.0
2,500-4,999	14.3	41.2	15.8
5,000-7,499	17.5	57.7	32.6
7,500-9,999	10.8	58.7	28.3
10,000-14,999	12.5	61.3	45.1
15,000-19,999	9.5	43.5	29.8
20,000-29,999	10.8	67.2	55.8
30,000-39,999	5.3	36.5	11.6
40,000-49,999	2.9	42.6	25.1
50,000-99,999	3.9	24.9	10.7
100,000 or more	0.8	12.0	–12.0
1980-1990 Population Growth			
Decrease of 10.0% or more	3.9	45.5	30.6
Decrease of 5.0-9.9%	12.0	54.7	34.9
Decrease of 0.1-4.9%	24.4	48.1	27.1
Increase of 0.0-4.9%	21.8	50.6	28.1
Increase of 5.0-9.9%	15.9	58.2	35.8
Increase of 10.0% or more	22.0	59.4	33.4
Percentage Poor School-Age Children, 1990			
0.0%	2.3	0.0	0.0
0.1%-5.9%	34.2	97.7	83.6
6.0-8.9%	16.1	42.1	20.1
9.0-12.4%	17.0	33.1	8.7
12.5-16.4%	15.1	26.4	3.0
16.5-23.9%	11.9	23.6	–15.6
24.0% or more	3.5	24.6	–20.6

Free Lunch		Free and Reduced-Price Lunch	
Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
48.7	7.1	52.1	14.0
57.4	4.6	59.8	10.1
40.1	–0.4	41.9	4.2
65.8	30.4	71.2	38.8
47.2	–12.1	45.7	–10.1
53.2	22.1	60.4	31.3
39.3	–6.3	43.1	1.3
47.3	14.1	54.3	27.8
37.2	–9.0	38.3	–3.3
36.7	–14.4	36.9	–8.0
24.2	–7.0	26.5	–2.9
5.1	5.1	4.7	–0.9
23.4	–0.9	31.8	10.0
63.0	7.0	67.5	11.0
39.3	–10.3	40.4	–4.9
47.9	20.4	51.2	27.7
43.0	12.0	47.2	20.1
60.6	11.3	64.6	19.4
0.0	0.0	0.0	0.0
81.4	22.2	90.2	37.6
41.9	2.7	44.8	8.9
41.0	6.8	41.7	10.6
22.6	–2.0	22.0	0.8
24.3	–10.4	23.4	–12.3
24.2	–16.3	24.5	–21.4

continued on next page

TABLE D-2 Continued

Category	Percent of Districts (N = 623)	Census Constant 1980 Share	
		Mean Absolute % Error	Mean Algebraic % Error
Change in Poverty Rates for Children, 1980-1990			
Decrease of 10.0% or more	4.5	132.1	129.3
Decrease of 5.0-9.9%	11.9	95.9	93.4
Decrease of 0.1-4.9%	46.1	55.8	50.2
Increase of 0.0-4.9%	29.2	23.9	–18.7
Increase of 5.0-9.9%	7.1	37.6	–37.1
Increase of 10.0% or more	1.3	59.1	–59.1
Percent of Population Black, 1990			
0.0-0.9%	15.1	29.9	9.9
1.0-4.9%	36.9	48.6	23.9
5.0-9.9%	34.7	64.5	42.0
10.0-24.9%	13.3	64.5	47.1
Percent of Population Hispanic, 1990			
0.0-0.9%	22.6	38.4	16.0
1.0-4.9%	49.3	48.8	28.4
5.0-9.9%	28.1	73.7	48.3

NOTES: The census constant 1980 share estimates are calculated as described in Chapter 7 (within-county shares method (2)). The school lunch estimates are formed by multiplying the 1990 census estimates of related children aged 5-17 in families in poverty for the county by the school district's share of the county's free (free and reduced-price) lunch participants. The mean unweighted abso-

Of the 623 districts in New York State that are in the census evaluation universe, 476 are in districts that had less than 20,000 total population in 1990. As shown in Table D-4, these 476 districts represent 76 percent of all districts in the evaluation universe for New York, but they contain only 35 percent of the poor children aged 5-17 in the census evaluation universe.

Tables D-5 to D-8 examine estimates of the number of districts and percentage of school-age children who are in poor families under alternative estimation methods in 1990. The census counts are the ratio-adjusted estimates of school-age children who are in poor families from the 1990 census. Census-based

Free Lunch		Free and Reduced-Price Lunch	
Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
101.8	48.1	103.3	51.8
64.5	47.5	73.4	57.2
53.4	11.5	57.0	20.5
33.8	–13.5	34.5	–8.6
26.4	–23.1	26.1	–21.5
44.2	–36.2	43.9	–42.0
33.3	5.3	34.8	9.8
46.7	17.2	50.0	22.0
54.0	0.2	57.5	8.2
57.8	0.9	64.2	11.7
41.3	13.8	45.3	19.2
40.8	9.9	44.1	18.4
68.5	–3.2	71.5	2.0

lute (algebraic) percentage error is the sum over all school districts of the absolute (algebraic or signed) difference between the estimate of poor school-age children from a model and the 1990 census estimate as a proportion of the census estimate for each district, divided by the number of districts.

estimates (within-county shares method (2) estimates) use the 1990 census counts of county school-age children who are in poor families and allocate these totals to school districts by the school district's share of county totals from the 1980 census. The model-based estimates (within-county shares method (1) estimates) use a similar approach, but with the county estimates of school-age children who are in poor families in 1989 produced from the Census Bureau's county model. The school lunch estimates are produced, as outlined above, by using 1990 county ratio-adjusted estimates of school-age children who are in poor families from the 1990 census and allocating them to constituent school districts by the share of

TABLE D-3 Mean Absolute and Algebraic Percentage Errors for Children Aged 5-17 in Families in Poverty, Various Methods, New York State School Districts in Evaluation Universe, 1990, Weighted by Children Aged 5-17 in Poverty, 1990 Census, in percent

		Census Constant 1980 Share	
Category	Percent of Districts (N = 623)	Mean Absolute % Error	Mean Algebraic % Error
Total	100.0	23.9	0.8
1990 School District Population			
Under 2,500	11.9	43.4	13.8
2,500-4,999	14.3	30.4	4.2
5,000-7,499	17.5	31.6	–0.1
7,500-9,999	10.8	32.5	–4.4
10,000-14,999	12.5	34.8	13.2
15,000-19,999	9.5	21.6	4.0
20,000-29,999	10.8	37.8	21.4
30,000-39,999	5.3	31.5	–2.0
40,000-49,999	2.9	33.6	9.4
50,000-99,999	3.9	18.3	–0.4
100,000 or more	0.8	10.4	–10.4
1980-1990 Population Growth			
Decrease of 10.0% or more	3.9	31.2	26.2
Decrease of 5.0-9.9%	12.0	13.7	–4.2
Decrease of 0.1-4.9%	24.4	20.8	–3.7
Increase of 0.0-4.9%	21.8	31.1	9.9
Increase of 5.0-9.9%	15.9	30.1	2.2
Increase of 10.0% or more	22.0	32.4	2.9
Percentage of Poor School-Age Children, 1990			
0.0%	2.3	0.0	0.0
0.1-5.9%	34.2	53.4	40.3
6.0-8.9%	16.1	34.0	9.6
9.0-12.4%	17.0	22.2	4.5
12.5-16.4%	15.1	22.7	–2.1
16.5-23.9%	11.9	19.0	–14.8
24.0% or more	3.5	11.3	–10.1

Free Lunch		Free and Reduced-Price Lunch	
Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
22.3	1.6	24.2	1.3
38.5	–7.4	39.2	–4.5
31.6	–11.2	31.8	–7.3
34.2	1.3	34.9	3.9
32.6	–4.6	30.0	–4.5
32.0	6.8	35.6	10.9
24.9	0.9	28.7	5.3
36.3	13.8	39.7	21.1
27.9	3.1	28.1	3.4
34.0	–6.8	35.6	–6.3
21.0	–1.5	24.0	–3.2
3.4	3.4	5.3	–3.0
9.8	2.2	24.2	7.7
10.2	–3.8	17.0	9.8
17.2	1.9	17.1	1.1
32.5	8.6	33.1	11.5
29.9	0.3	30.6	3.3
39.0	2.0	38.0	4.6
0.0	0.0	0.0	0.0
47.1	–5.3	52.1	8.7
34.3	2.8	36.7	8.6
36.3	19.5	38.5	23.5
19.6	–1.6	20.0	–2.3
18.2	1.4	16.2	–3.3
5.4	–0.6	9.1	–7.8

continued on next page

TABLE D-3 Continued

Category	Percent of Districts (N = 623)	Census Constant 1980 Share: Mean Absolute % Error	Census Constant 1980 Share: Mean Algebraic % Error
Change in Poverty Rates for Children, 1980-1990			
Decrease of 10.0% or more	4.5	79.3	75.3
Decrease of 5.0-9.9%	11.9	47.0	38.1
Decrease of 0.1-4.9%	46.1	32.1	23.1
Increase of 0.0-4.9%	29.2	19.0	–15.0
Increase of 5.0-9.9%	7.1	16.9	–15.7
Increase of 10.0% or more	1.3	10.6	–10.6
Percent of Population Black, 1990			
0.0-0.9%	15.1	21.4	–0.5
1.0-4.9%	36.9	24.4	0.5
5.0-9.9%	34.7	30.1	2.4
10.0-24.9%	13.3	16.9	–0.6
Percent of Population Hispanic, 1990			
0.0-0.9%	22.6	22.7	0.7
1.0-4.9%	49.3	20.1	–0.5
5.0-9.9%	28.1	34.8	4.1

NOTE: See notes to Table D-2.

that school district's free (or free and reduced-price) school lunch eligibles relative to the county total.

Tables D-5 and D-6 provide estimates for the two-tier concentration grant eligibility for districts with total population (from the 1990 census) of less than 20,000. That is, districts must be in counties where at least 15 percent (or 6,500) of the school-age children are poor and in a district that also meets this criterion.[4] If we take the census counts as our measure of "truth," then employing school

[4]In Tables D-5 and D-6, county eligibility is determined by the county counts from the 1990 census. Within each of these eligible counties, the alternative methods listed are used to determine school district eligibility.

Free Lunch		Free and Reduced-Price Lunch	
Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
64.9	19.5	68.2	23.5
41.4	31.7	43.1	32.4
36.0	7.1	37.9	12.0
20.9	–4.8	21.1	–3.6
8.5	–5.2	12.0	–10.5
5.5	–1.2	7.8	–7.7
22.4	–2.6	23.8	–0.9
23.6	0.5	24.6	0.0
32.0	4.6	32.1	4.1
9.8	0.0	14.5	–0.3
23.5	0.0	25.5	0.7
15.0	0.5	18.1	0.0
40.8	5.4	39.3	5.2

lunch data will likely overstate eligibility. As shown in Table D-5, roughly 50 percent more districts and school-age children are estimated to be eligible with free school lunch data than with the census counts. This problem is further magnified when the free and reduced-price lunch counts are employed. Table D-6 illustrates where each method errs relative to the eligibility categorization of the census counts: as might be expected, the school lunch estimates produce a substantial number of false positives.

Tables D-7 and D-8 provide a similar analysis for direct allocations. Now districts must only meet the single criterion that the district has at least 15 percent of its school-age children who are poor (or at least a total of 6,500). These estimates also show the effect of imposing county controls on the use of school

TABLE D-4 New York State Districts in Evaluation Universe Above and Below the 20,000 Population Threshold for Pooling Allocations (N = 623)

	Districts		Poor Children Aged 5-17	
Category	Number	Percent	Number	Percent
School District Total Population				
Less than 20,000	476	76.4	61,236	35.0
At least 20,000	147	23.6	113,556	65.0

TABLE D-5 Concentration Grant Eligibility at County and School District Level, Various Methods for New York State Districts in Evaluation Universe with Less than 20,000 Population, 1990 (N = 476)

	Districts		Poor Children Aged 5-17	
Method	Number	Percent	Number	Percent
Census Counts	76	16.0	16,689	27.3
Census-based Estimates	78	16.4	14,162	23.1
Model-based Estimates	76	16.0	14,134	23.1
Free Lunch[a]	112	23.5	21,662	35.4
Free and Reduced-price Lunch[a]	136	28.6	24,515	40.0

NOTES: Cell entries are for school districts and poor school-age children that would be eligible for concentration grants according to various methods (see text) under the current two-stage allocation process (i.e., both county and school district have more than 6,500 or more than 15% poor school-age children). The total number of poor school-age children in districts with less than 20,000 population is 61,236.

[a]Some school districts (54 or 11.3%) did not report school lunch data.

lunch estimates. (The county controls are equivalent to the estimates produced by equation 1, above.) The school lunch estimates without controls greatly overstate concentration grant eligibility. Imposing county controls substantially improves the accuracy of these estimates.

Table D-9 shows mean algebraic and absolute percentage errors for the various estimation methods. Here the school lunch estimates have either been

TABLE D-6 Concentration Grant Eligibility at County and School District Level, Census Counts Compared to Various Other Methods for New York State Districts in Evaluation Universe with Less than 20,000 Population, 1990 (N = 476)

	Census Not Eligible				Census Eligible			
	Estimate Not Eligible		Estimate Eligible		Estimate Not Eligible		Estimate Eligible	
Method	% Districts	% Poor Children Aged 5-17	% Districts	% Poor Children Aged 5-17	% Districts	% Poor Children Aged 5-17	% Districts	% Poor Children Aged 5-17
Census-based Estimates	78.6	68.0	5.5	4.8	5.0	8.9	10.9	18.4
Model-based Estimates	78.8	68.1	5.3	4.6	5.3	8.8	10.7	18.4
Free Lunch[a]	75.4	63.6	8.6	9.2	1.1	1.0	14.9	26.2
Free and Reduced-price Lunch[a]	71.2	59.8	12.8	13.0	0.2	0.2	15.8	27.1

NOTE: See notes to Table D-5.
[a]Some school districts (54 or 11.3%) did not report school lunch data.

TABLE D-7 Concentration Grant Eligibility at School District Level, Various Methods for New York State Districts in Evaluation Universe with Less than 20,000 Population, 1990 (N = 476)

	Districts		Poor Children Aged 5-17	
Method	Number	Percent	Number	Percent
Census Counts	115	24.2	25,343	41.4
Census-based Estimates	114	24.0	19,596	32.0
Model-based Estimates	109	22.9	18,285	29.9
Free Lunch[a]	214	45.0	39,222	64.1
Free and Reduced-price Lunch[a]	294	61.8	48,835	79.8
Free Lunch with Controls[a,b]	124	26.1	25,024	40.9
Free and Reduced-price Lunch with Controls[a,b]	127	26.7	24,045	39.3

NOTES: Cell entries are for school districts and poor school-age children that would be eligible for concentration grants according to various methods (see text) under a direct allocation process (i.e., the school district has more than 6,500 or more than 15% poor school-age children). The total number of poor school-age children in districts with less than 20,000 population is 61,236.

[a]Some school districts (54 or 11.3%) did not report school lunch data.

[b]Controls are imposed at the county level so that number of poor children and number of children in the school district must sum to county census counts for 1990.

controlled to the statewide total of school-age children living in poor families for the 476 districts with populations of less than 20,000 or to a similar county total. With these controls in place, each of the methods has roughly the same algebraic and absolute percentage errors. This result is interesting as the school lunch estimates with county controls had the potential to be either better or worse than the estimates with state controls. We would in general expect them to be better as there is a tighter level of control imposed. It is possible that they are worse as a result of lack of precision that occurs when school districts cross county boundaries and school lunch data are coded to the county where the district office is located.

PROBLEMS WITH THIS APPROACH

Using school lunch data to estimate the number of poor children for each school district has several potential problems, based on the experience in New York State. As has been widely acknowledged:

TABLE D-8 Concentration Grant Eligibility at School District Level, Census Counts Compared to Various Other Methods for New York State Districts in Evaluation Universe with Less than 20,000 Population, 1990 (N = 476)

	Census Not Eligible				Census Eligible			
	Estimate Not Eligible		Estimate Eligible		Estimate Not Eligible		Estimate Eligible	
Method	% Districts	% Poor Children Aged 5-17	% Districts	% Poor Children Aged 5-17	% Districts	% Poor Children Aged 5-17	% Districts	% Poor Children Aged 5-17
Census-based Estimates	65.3	49.5	10.5	9.1	10.7	18.5	13.5	22.9
Model-based Estimates	66.6	52.1	9.2	6.5	10.5	18.1	13.7	23.3
Free Lunch[a]	52.5	34.0	23.3	24.7	2.5	2.0	21.6	39.4
Free and Reduced-price Lunch[a]	37.0	19.7	38.9	38.9	1.3	0.6	22.9	40.8
Free Lunch with Controls[a,b]	66.8	50.0	9.0	8.7	7.1	9.2	17.0	32.2
Free and Reduced-price Lunch with Controls[a,b]	65.6	49.8	10.3	8.9	7.8	11.0	16.4	30.4

NOTE: See notes to Table D-7.

[a]Some school districts (54 or 11.3%) did not report school lunch data.

[b]Controls are imposed at the county level so that number of poor children and number of children in the school district must sum to county census counts for 1990.

TABLE D-9 Mean Absolute and Algebraic Percentage Errors for Children Aged 5-17 in Families in Poverty, Various Methods, Unweighted and Weighted, New York State School Districts in Evaluation Universe with Less than 20,000 Population, 1990 (N = 476)

	Unweighted		Weighted by Poor School-Age Children	
Method	Mean Absolute % Error	Mean Algebraic % Error	Mean Absolute % Error	Mean Algebraic % Error
Census-based Estimates	54.9	30.8	30.6	4.3
Model-based Estimates	54.3	27.2	31.5	0.1
Free Lunch[a]	52.9	9.8	32.3	–0.1
Free and Reduced-price Lunch[a]	55.1	13.4	32.6	–0.1
Free Lunch with Controls[b]	54.6	12.6	31.7	–0.1
Free and Reduced-price Lunch with Controls[b]	55.7	13.3	31.7	–0.1

NOTES: See text for the calculation of poor school-age children by each method. The mean unweighted algebraic percentage error is the sum over all school districts of the algebraic difference between the estimate of poor school-age children from a model and the 1990 census estimate as a proportion of the census estimate for each district, divided by the number of districts. The weighted mean weights each difference by the census number of poor school-age children in the district.

[a]Controls are imposed at the state level so that the number of poor children and number of children in the school district must sum to the state census count for 1990 for districts with less than 20,000 population.

[b]Controls are imposed at the county level so that number of poor children and number of children in the school district must sum to county census counts for 1990 for districts with less than 20,000 population.

- Participants in the National School Lunch Program are not the target population of Title I:
  - — There are differences in eligibility between Title I and school lunch.
  - — There are differences in reporting geography: Title I counts residents, school lunch counts by location of the school the child attends.
- Not all eligible students apply for the school lunch program, and application rates appear to be uneven across schools.
- Some schools choose not to participate in the school lunch program.

Other difficulties include:

- New York State has a number of regional (groups of counties) educational authorities with students, and they participate in the school lunch program; how to allocate these students is an issue.
- In New York State, the school lunch program is administered separately from most other programs, which can make use of the administrative data difficult (e.g., schools sometimes have separate identification numbers, which makes matching to other data very time consuming).

APPENDIX

E

Special Case: Estimates for Puerto Rico

Puerto Rico is included in the Title I fund allocations. Since the commonwealth has no administrative subdivisions, the Department of Education treats it as a single unit (equivalent to a U.S. county and coterminous school district) for the allocation of these funds. In order to incorporate Puerto Rico in the fiscal 1997 fund allocation for school year 1997-1998, estimates of its number and proportion of related children aged 5-17 living in poverty were needed for 1993.

If the allocations for school year 1997-1998 had been based on 1990 census estimates (which the panel did not recommend), the estimates for Puerto Rico could have been obtained straightforwardly from the commonwealth's 1990 decennial census. From that census it is estimated that Puerto Rico had about 558,000 poor related children aged 5-17 in 1989, 66.4 percent of all related children in this age range. However, the panel recommended that the 1997-1998 allocations be based in part on estimates of the number and proportion of school-age children in poverty in 1993, and it was not straightforward to develop such estimates for Puerto Rico.

The Puerto Rico Bureau of Labor Statistics conducts a periodic labor force survey, but that survey does not collect CPS-type income information on a regular basis. In addition, the specific model-based estimation procedures developed by the Census Bureau for U.S. states and counties cannot be applied to Puerto Rico since they are based on tax return and food stamp participation data for which there are no precise equivalents for Puerto Rico.

The only available data source for updating estimates of poor school-age children in Puerto Rico was an experimental March 1995 income survey modeled after the CPS March Income Supplement. The Census Bureau based its 1993 estimates of poor school-age children on data from this survey, together with data for Puerto Rico from the decennial census and updated population estimates.

The derivation of the estimates of poor school-age children in Puerto Rico in 1993 from these data sources required a number of adjustments, for several reasons: (1) the March 1995 experimental survey did not collect information on the ages of family members under 18 (so that related children aged 5-17 could not be identified among those aged under 18); (2) the updated Puerto Rico population estimates are for all children in the resident population, not for related children only; and (3) the survey, which was conducted in 1995, obtained information on 1994 income, not 1993 income. In making the adjustments, the Census Bureau assumed that certain relationships observed in 1990 census data still applied and that the change in the number of Puerto Rico school-age children in poverty between 1989 and 1994 was linear.

The panel did not have any data with which to test the validity of these assumptions. It had only limited information about the sample design, sampling and nonsampling errors, response rates, and other features of the experimental survey. The sample size of about 3,200 households should be large enough to provide a direct estimate of the number of poor school-age children with adequate precision. However, only limited information was available about other key aspects of data quality, including response rates for households to the income questions and the editing or imputation procedures used.

The Census Bureau computed 1995 estimates for Puerto Rico from data collected in the 1996 Puerto Rican Family Income Survey that was conducted in the commonwealth in February-March 1997. (The survey is planned to be conducted at regular intervals in the future.) Several adjustments had to be made to produce the estimates of school-age children in poverty in 1995. The approach used was similar to that used to compute 1993 estimates of poor school-age children. Additional information was obtained from Puerto Rico about the quality of the income survey that, in general, supported the use of the survey data to develop 1995 estimates of the number of poor school-age children for Puerto Rico (see Santas and Waddington, 1999). Consequently, the panel recommended that the 1995 estimates for Puerto Rico be used in the direct Title I allocations for the 1999-2000 school year.

The Puerto Rico Family Income Survey will presumably be the basis of updated estimates of poor school-age children in Puerto Rico for 1997 and later years. Through cooperative work with Puerto Rico, the Census Bureau should continue its evaluations of the quality of the estimates and their comparability with the model-based estimates for U.S. counties to determine if there are ways in which the data and estimation procedures for Puerto Rico can be improved for use in Title I allocations.

References and Bibliography

NOTE: Many of the cited reports and papers by Census Bureau staff are available on the Census Bureau's web site: http://www.census.gov. Other reports related to small-area income and poverty estimates may be obtained by contacting: David Waddington, Division of Housing and Household Economic Statistics, U.S. Bureau of the Census, Washington, D.C. 20233.

Papers that were prepared for the panel—marked by an asterisk (*)—are available from the Committee on National Statistics, 2101 Constitution Ave. NW, Washington, D.C. 20418.

Abelson, R.P., and J.W. Tukey
1963 Efficient utilization of nonnumerical information in quantitative analysis: General theory and the case of simple order. *Annals of Mathematical Statistics* 34:1347-1369.
Alexander, C.H.
1998 Recent Developments in the American Community Survey. Paper prepared for presentation at the Joint Statistical Meetings, Dallas. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (August).
Alexander, C.H., S. Dahl, and L. Weidman
1997 Making Estimates from the American Community Survey. Paper prepared for presentation at the Joint Statistical Meetings, Anaheim, Calif. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (August).
Bell, W.R.
1997a A County CPS Model with "Census Residuals." Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (July)
1997b Regression Diagnostics for Models for County Poverty Estimates–Complete. Statistical Research Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (September).
1999 Accounting for uncertainty about variances in small area estimation. Proceedings of the Section on Survey Research Methods. Alexandria, Va.: American Statistical Association.

Bell, W.R., P.W. Cardiff, R.C. Fisher, E.R. Miller, M. Kramer, P.M. Siegel, A. Strand, and D.G. Waddington

2000 Poverty Estimation for U.S. States, Counties, and School Districts. Paper prepared for the Joint Statistical Meetings, Indianapolis. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (August).

Belsley, D.A., E. Kuh, and R.E. Welsch

1980 *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Betson, D.M.*

1999 Replication of Jim [Wyckoff's] New York Lunch Analysis Using Indiana Data. Prepared for Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics, National Research Council. Department of Economics, University of Notre Dame, Ind.

Cardiff, P.

1998 IRS Tax Return Data and Poor Children. Draft. Housing and Household Economic Statistics Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (September).

Coder, J.F., R.C. Fisher, and P.M. Siegel

1996 Making Model-Based Estimates of the Number of Related Persons Age 5-17 in Poverty for all U.S. Counties, 1994. Housing and Household Economic Statistics Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (November).

Davis, S.

1994 *Evaluation of Post Censal County Estimates for the 1980's*. Population Division Working Paper Series, Number 5. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Fay, R.E.

1996 Methodology for Poverty Estimation for 1993. Section 2. Demographic Statistical Methods Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (November).

1997a Comparisons of CPS and Census Poverty Levels. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (October).

1997b Notes on Direct Variance Estimates for Counties. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (September).

Fay, R.E., and R.A. Herriot

1979 Estimates of income for small places: An empirical Bayes application of James-Stein procedures to census data. *Journal of the American Statistical Association* 78:269-277.

Fay, R.E., C.T. Nelson, and L. Litow

1993 Estimation of median income for 4-person families by state. Pp. 9-1 to 9-17 in *Indirect Estimators in Federal Programs*. Statistical Policy Working Paper 21, Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget. Washington, D.C.: U.S. Government Printing Office.

Fay, R.E., and G.F. Train

1995 Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties. *Proceedings of the Section on Government Statistics*. Alexandria, Va:. American Statistical Association.

1997 Small domain methodology for estimating income and poverty characteristics for states in 1993. *Proceedings of the Social Statistics Section*. Alexandria, Va.: American Statistical Association.

Fisher, G.M.

1992 Poverty guidelines for 1992. *Social Security Bulletin* 55(1)(Spring):43-46.

Fisher, R.
1997 Methods used for small area poverty and income estimation. *Proceedings of the Social Statistics Section*. Alexandria, Va.: American Statistical Association.

Fisher, R., and J. Asher
1999a Alternate CPS Sampling Variance Structures for Constrained and Unconstrained County Models. Internal Technical Report Series #1. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (September).
1999b Bayesian Hierarchical Modeling of U.S. County Poverty Rates. Paper presented at Case Studies in Bayesian Statistics Workshop. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (September).

Fuller, W.A., and J.J. Goyeneche
1998 Estimation of the State Variance Component. Working Paper. Statistical Laboratory, Iowa State University, Ames. (June).

Ghosh, M. and J.N.K. Rao
1994 Small-area estimation: An appraisal. *Statistical Science* 8 (1):55-93.

Long, J.
1993 *Post Censal Population Estimates: States, Counties, and Places*. Population Division Working Paper Series, Number 3. Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Midwest Research Institute*
1999 Uses of Small-Area Poverty Estimates. Final Report. By L. Butcher and N. Dunton. Paper prepared for the Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics, National Research Council. Midwest Research Institute, Kansas City, Mo.

Moskowitz, J., S. Stullich, and B. Deng
1993 *Targeting, Formula, and Resource Allocation Issues: Focusing Federal Funds Where the Needs are Greatest*. Washington, D.C.: U.S. Department of Education.

National Research Council
1985 *The Bicentennial Census: New Directions for Methodology in 1990*. Panel on Decennial Census Methodology, C.F. Citro and M.L. Cohen, eds., Committee on National Statistics. Washington, D.C.: National Academy Press
1993 *The Future of the Survey of Income and Program Participation*. Panel to Evaluate the Survey of Income and Program Participation, C.F. Citro and G. Kalton, eds., Committee on National Statistics. Washington, D.C.: National Academy Press.
1995a *Measuring Poverty: A New Approach*. Panel on Poverty and Family Assistance: Concepts, Information Needs, and Measurement Methods, C.F. Citro and R.T. Michael, eds., Committee on National Statistics. Washington, D.C.: National Academy Press.
1995b *Modernizing the U.S. Census.* Panel on Census Requirements in the Year 2000 and Beyond, B. Edmonston and C. Schultze, eds., Committee on National Statistics. Washington, D.C.: National Academy Press.
1997 *Small-Area Estimates of Children in Poverty, Interim Report 1, Evaluation of 1993 County Estimates for Title I Allocations*. Panel on Estimates of Poverty for Small Geographic Areas, C.F. Citro, M.L. Cohen, G. Kalton, and K.K. West, eds., Committee on National Statistics. Washington, D.C.: National Academy of Press.
1998 *Small-Area Estimates of Children in Poverty, Interim Report 2, Evaluation of Revised 1993 County Estimates for Title I Allocations*. Panel on Estimates of Poverty for Small Geographic Areas, C.F. Citro, M.L. Cohen, and G. Kalton, eds., Committee on National Statistics. Washington, D.C.: National Academy of Press.

1999 *Small-Area Estimates of Children in Poverty, Interim Report 3, Evaluation of 1995 County and School District Estimates for Title I Allocations*. Panel on Estimates of Poverty for Small Geographic Areas, C.F. Citro and G. Kalton, eds., Committee on National Statistics. Washington, D.C.: National Academy of Press.

2000 *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond*. Panel on Estimates of Poverty for Small Geographic Areas, C.F. Citro and G. Kalton, eds., Committee on National Statistics. Washington, D.C.: National Academy of Press.

Otto, M.C., and W.R. Bell

1995 Sampling error modeling of poverty and income statistics for states. *American Statistical Association Proceedings of the Section on Government Statistics*. Alexandria, Va.: American Statistical Association.

Platek, R., J.N.K. Rao, C.E. Särndal, and M.P. Singh, eds.

1987 *Small Area Statistics*. New York: John Wiley and Sons.

Prasad, N.G.N., and J.N.K. Rao

1990 The estimation of mean squared errors of small area estimators. *Journal of the American Statistical Association* 78:47-59.

Rao, J.N.K.

1999 Small-Area Estimation: Updates with Appraisal. Unpublished paper. Carleton University, Ottawa.

Reed, J.M.

1996 Puerto Rico Estimates and Projections Based on Demographic Component Method: 1990-1995. International Programs Center, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (November).

Robinson, J.G., B. Ahmed, P. Das Gupta, and K. Woodrow

1993 Estimation of population coverage in the 1990 U.S. based on demographic analysis. *Journal of the American Statistical Association* 88(423):1061-1079.

Santos, R., and D. Waddington*

1999 Obtaining Estimates of Child Poverty in Puerto Rico: An Overview of Current Practices and Recommendations for Improvement. Report prepared for the Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics, National Research Council. National Opinion Research Center, Chicago, Ill., and Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

Siegel, P.

1997 Background on Poverty Estimates for School Districts [Tables]. Housing and Household Economic Statistics Division, Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (May).

Siegel, P., and R. Fisher

1998 Sampling Variances of the 1990 Census Estimates of Numbers of Related Children Ages 5-17 in Families in Poverty for School Districts. Draft. Housing and Household Economic Statistics Division, Bureau of the Census, Washington, D.C. (September).

Sink, L.

1996 *Estimates of the Population of Counties by Age, Sex, Race, and Hispanic Origin: 1990-1994*. Release PE-48 (Methodology). Bureau of the Census. Washington, D.C.: U.S. Department of Commerce.

Spencer, B.D., and Che-Fu Lee

1980 Postcensal population estimation methods of the Census Bureau. Pp. 131-187 in *Estimating Population and Income of Small Areas*. Panel on Small-Area Estimates of Population and Income, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.

Thiel, H.

1971 *Principles of Econometrics*. New York: John Wiley and Sons.

U.S. Census Bureau

1987 *State Population and Household Estimates, with Age, Sex, and Components of Change: 1981-86*. Current Population Reports, Series P-25, No. 1010. Washington, D.C.: U.S. Department of Commerce.

1990 *Money Income and Poverty Status in the United States, 1989.* Current Population Reports, Series P60-168. Washington, D.C. U.S. Department of Commerce.

1993 *Poverty in the United States: 1992*. Current Population Report, Series P60-185. Washington, D.C.: U.S. Department of Commerce.

1995a *Income, Poverty, and Validation of Noncash Benefits: 1993*. Current Population Reports, Series P60-188. Washington, DC: U.S. Department of Commerce.

1995b *Subnational Estimates of Total Population by the Tax Return Method*. Population Estimates Branch, Population Division. Washington, D.C.: U.S. Department of Commerce.

1996 *Poverty in the United States: 1995.* Current Population Reports, Series P60-194. Washington, D.C.: U.S. Department of Commerce.

1998a 1995 SAIPE State and County Models. Draft documentation. By P. Cardiff, R. Fisher, P. Siegel, A. Strand, D. Waddington, W.R. Bell, and M. Kramer, with assistance from R.E. Fay and G. Train. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (October).

1998b Tables 1-24 [comparing 1990 census estimates and estimates produced by several methods for poor school-age children, total school-age children, and total population in school districts]. Bureau of the Census, U.S. Department of Commerce, Washington, D.C. (November).

2000 *Poverty in the United States: 1999.* Current Population Reports, Series P-60-210. Washington, D.C.: U.S. Department of Commerce.

U.S. Census Bureau and Bureau of Labor Statistics

2000 *Current Population Survey Design and Methodology*. Technical paper 63. Washington, D.C.: U.S. Department of Commerce and U.S. Department of Labor.

U.S. Office of Management and Budget

1993 *Indirect Estimators in Federal Programs*. Statistical Policy Working Paper 21. Subcommittee on Small-Area Estimation, Federal Committee on Statistical Methodology, Statistical Policy Office, Office of Information and Regulatory Affairs. Washington, D.C.: U.S. Government Printing Office.

Voss, P.R., R. Gibson, and K. Morgen*

1997 Assessment of the Initial 1993 Estimates of School-Age Children in Poverty. Report prepared for the Panel on Estimates of Poverty for Small Geographic Areas, Committee on National Statistics, National Research Council. Department of Rural Sociology, University of Wisconsin-Madison.

Biographical Sketches of Panel Members and Staff

GRAHAM KALTON (*Chair*) is a senior statistician and senior vice president of Westat. He is also a research professor in the Joint Program in Survey Methodology at the University of Maryland. Previously, he was a research scientist in the Survey Research Center and a professor of biostatistics and statistics at the University of Michigan, professor of social statistics at the University of Southampton, and reader in social statistics at the London School of Economics. His research interests are in survey sampling and general survey methodology. He is a fellow of the American Statistical Association and of the American Association for the Advancement of Science. He has served as president of the International Association of Survey Statisticians and president of the Washington Statistical Society. He is a past member of the Committee on National Statistics and has served as chair or a member of several of its panels. He received a B.Sc. degree in economics and an M.Sc. degree in statistics from the University of London and a Ph.D. degree in survey methodology from the University of Southampton.

DAVID M. BETSON is associate professor of economics at the University of Notre Dame. His previous positions have been as a visiting scholar at the Joint Center for Poverty Research of the University of Chicago and Northwestern University, a research associate at the Institute for Research on Poverty at the University of Wisconsin, and an economist in the Office of the Assistant Secretary for Planning and Evaluation in the U.S. Department of Health, Education, and Welfare. His research examines the effects of governments on the distribution of economic well-being with special reference to the measurement of poverty and the analysis of child support policy. He received a Ph.D. degree in economics from the University of Wisconsin–Madison.

CONSTANCE F. CITRO is a senior program officer for the Committee on National Statistics. She is a former vice president and deputy director of Mathematica Policy Research, Inc., and was an American Statistical Association/ National Science Foundation research fellow at the U.S. Census Bureau. For the committee, she has served as study director for numerous projects, including the Panel on Poverty and Family Assistance, the Panel to Evaluate the Survey of Income and Program Participation, the Panel to Evaluate Microsimulation Models for Social Welfare Programs, and the Panel on Decennial Census Methodology. Her research has focused on the quality and accessibility of large, complex microdata files, as well as analysis related to income and poverty measurement. She is a fellow of the American Statistical Association. She received a B.A. degree from the University of Rochester and M.A. and Ph.D. degrees in political science from Yale University.

MICHAEL L. COHEN is a senior program officer for the Committee on National Statistics, currently serving as the study director for the Panel on Research on Future Census Methods and assisting the Panel on Estimates of Poverty for Small Geographic Areas. He previously directed the Panel on Statistical Methods for Testing and Evaluating Defense Systems. Formerly, he was a mathematical statistician at the Energy Information Administration, an assistant professor in the School of Public Affairs at the University of Maryland, and a visiting lecturer at the Department of Statistics, Princeton University. His general area of research is the use of statistics in public policy, with particular interest in census undercount, model validation, and robust estimation. He received a B.S. degree in mathematics from the University of Michigan and M.S. and Ph.D. degrees in statistics from Stanford University.

NANCY E. DUNTON is director of the Center for Health and Social Research at the Midwest Research Institute. Formerly, she was a senior research scientist with the New York State Department of Social Services and the New York State Council on Children and Families. Her research focuses on social service policy issues, with a special emphasis on welfare reform and children's well-being. She received a Ph.D. degree in sociology from the University of Wisconsin-Madison.

WAYNE A. FULLER is distinguished professor in the Department of Statistics and Economics at Iowa State University. He is a fellow of the American Statistical Association, the Econometric Society, and the Institute of Mathematical Statistics. He is the author of *Introduction to Statistical Time Series* and *Measurement Error Models* and has an active research program in survey sampling. He has held offices in national and international statistical organizations and has previously served on National Research Council panels. He received B.S. and M.S. degrees and a Ph.D. degree in statistical theory and methods from Iowa State University.

THOMAS B. JABINE is a statistical consultant who specializes in the areas of sampling, survey research methods, and statistical policy. He was formerly a statistical policy expert for the Energy Information Administration, chief mathematical statistician for the Social Security Administration, and chief of the Statistical Research Division of the U.S. Census Bureau. He is a fellow of the American Statistical Association and a member of the International Statistical Institute. He has a B.S. degree in mathematics and an M.S. degree in economics and science from the Massachusetts Institute of Technology.

SYLVIA T. JOHNSON is professor of research methodology and statistics in the School of Education at Howard University, where she is also editor-in-chief of the *Journal of Negro Education*. She is a principal investigator at the Center for Research on the Education of Students Placed at-Risk (CRESPAR), a joint activity of Howard University and Johns Hopkins University, and chair of the Design and Analysis Committee of the National Assessment of Educational Progress (NAEP). She has served on the faculties of Roosevelt, Trenton (NJ) State, and Western Illinois universities and Augustana and Chicago City Colleges; as a visiting scholar at the Educational Testing Service; as a member of the Technical Advisory Committee for the National Adult Literacy Survey and the Graduate Record Examination; and as a trustee of the College Board. She is a fellow of the American Psychological Association and chaired its Division 15 Committee on Members and Fellows and has served on the board of directors of the National Council on Measurement in Education. Her research interests include test fairness and use and the academic achievement of African American students. Dr. Johnson received a B.S. degree in mathematics from Howard University and a Ph.D. degree in educational measurement and statistics from the University of Iowa.

THOMAS A. LOUIS is professor in the division of biostatistics at the University of Minnesota School of Public Health and professor of statistics in the School of Statistics. From 1987 to 1999 he was head of Biostatistics. In August 2000, he became a senior statistician at RAND. He is a member of the Committee on National Statistics, on the board of the Medical Follow-up Agency of the Institute of Medicine, and on the executive committee of the National Institute of Statistical Sciences. He is an elected member of the International Statistics Institute and a fellow of the American Statistical Association and the American Association for the Advancement of Science. His research interests focus on public health issues, including disease mapping and the health effects of airborne particulates, and on the development of statistical procedures, including Bayesian methods and the analysis of observational studies. He received a B.A. degree from Dartmouth College and a Ph.D. degree in mathematical statistics from Columbia University.

SALLY C. MORTON is the Rand Chair in Statistics and head of the Statistics Group at RAND in Santa Monica, California. She is a lecturer in the School of Public Health at the University of California–Los Angeles. Previously, she taught at the RAND Graduate School of Public Policy Studies and the School of Business of the University of Southern California, was a visiting scholar at the Centre for Mathematical Analysis at Australian National University and the University of Southampton, and worked at the Stanford Linear Accelerator Center. She serves as an editor of *Statistical Science* and as an associate editor for the *Journal of the American Statistical Association*, is past chair of the association's Section on Statistical Graphics, and is a member of the Caucus for Women in Statistics and the Institute of Mathematical Statistics. Her health policy research concerns homelessness, severe mental illness, and outcomes research and quality of care in the areas of childbirth and AIDS. Her methodological research concentrates on meta-analysis, nonparametric regression, and the sampling of vulnerable populations. She received a Ph.D. degree in Statistics from Stanford University.

JEFFREY S. PASSEL is a principal research associate at the Urban Institute. Previously, he was assistant division chief for estimates and projections in the Population Division of the U.S. Census Bureau, and he also directed the research on demographic methods for measuring census undercount. His research interests include the demography of immigration, particularly the measurement of illegal immigration; the effects and integration of immigrants into American society; and measuring and defining racial and ethnic groups in the United States. He is a member of a number of professional societies and has served in various capacities in the Population Association of America, the American Statistical Association, and the American Association for the Advancement of Science. He is a fellow of the American Association for the Advancement of Science. He received a B.S. degree in mathematics from the Massachusetts Institute of Technology, an M.A. degree in sociology from the University of Texas at Austin, and a Ph.D. degree in social relations from the Johns Hopkins University.

J.N.K. RAO is professor of statistics at Carleton University in Ottawa, Canada, and a consultant to Statistics Canada. Formerly, he was a professor at the University of Manitoba and Texas A&M University. His research interests include survey sampling theory and methods, particularly small-area estimation. He is a fellow of the Royal Society of Canada, the American Statistical Association, and the Institute of Mathematical Statistics. He received a Ph.D. degree in statistics from Iowa State University.

ALLEN L. SCHIRM is senior fellow at Mathematica Policy Research, Inc. Formerly, he was Andrew W. Mellon assistant research scientist and assistant professor at the University of Michigan. His principal research interests include small-area estimation and sample and evaluation design, with application to studies of child well-being and welfare, food and nutrition, and education policy. He is a member of the Committee on National Statistics Panel on Research on Future Census Methods and the American Statistical Association/Section on Survey Research Methods Working Group on the Technical Aspects of the Survey of Income and Program Participation. He received an A.B. degree in statistics from Princeton University and a Ph.D. degree in economics from the University of Pennsylvania.

MICHELE VER PLOEG, a research associate for the panel, is also serving as study director for the Panel on Data and Methods for Measuring the Effects of Changes in Social Welfare Programs of the Committee on National Statistics. Her research interests include the effects of social policies on families and children, the outcomes of children who experience poverty and changes in family composition, and individuals' education attainment choices. She received a B.A. degree in economics from Central College and M.S. and Ph.D. degrees in consumer economics and housing from Cornell University.

PAUL R. VOSS is professor of rural sociology at the University of Wisconsin–Madison and is affiliated with the Wisconsin Institute for Research on Poverty. For the past 21 years he has been affiliated with the Wisconsin Applied Population Laboratory and is currently its director. His research involves modeling small-area population change for purposes of population estimation and projection, and he also has studied and written about the demographic composition of small-area migration streams. He is a member of the U.S. Department of Commerce's Advisory Committee for the 2000 Census, as well as the Census Bureau's Advisory Committee of Professional Associations. He received a Ph.D. degree in sociology (demography) from the University of Michigan.

JAMES H. WYCKOFF is acting dean in the Graduate School of Public Affairs at the University at Albany, State University of New York, and associate professor of public administration, public policy, and economics. His research focuses on the economics of education and state and local public economics. He was an American Statistical Association Fellow at the U.S. Census Bureau. He received a Ph.D. degree in economics from the University of North Carolina at Chapel Hill.

ALAN M. ZASLAVSKY is associate professor of statistics in the Department of Health Care Policy at Harvard Medical School. He was formerly on the faculty of the Department of Statistics at Harvard. His research interests include measurement of quality in health care, census methodology, estimation and correction of census undercount, small-area estimation, microsimulation, design and analysis of surveys, and Bayesian methods. He has served on two other panels of the Committee on National Statistics concerned with planning for the 2000 census. He received a Ph.D. degree in applied mathematics from the Massachusetts Institute of Technology.

MEYER ZITTER, a consultant for the panel, is an independent demographic consultant. Formerly, he was chief of the U.S. Census Bureau's Population Division and also served as assistant director for international programs. He is a fellow of the American Statistical Association and a member of the International Statistical Institute and the International Union for the Scientific Study of Population. He has a B.B.A. degree from City College of New York.